THREE ESSAYS IN APPLIED MICROECONOMICS

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA

BY

NAPAT JATUSRIPITAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PAUL GLEWWE, GERARD MCCULLOUGH

March 2018

For my son, Nash.

Without you I would still be writing this dissertation

Introduction

This dissertation consists of three independent chapters that study how to improve public policies and reduce the level of social injustice through the lens of microeconomics and the innovative use of new data sets. In the first chapter, I test the impact of neighborhood heterogeneity on the private contribution of local public goods. Using a panel data set containing over two million non-emergency service requests and detailed census-tract level data on socioeconomic characteristics from the American Community Survey, I find that, contrary to the prevailing view in the literature, racial and linguistic heterogeneity have little to no negative effect on private voluntary contributions to local public goods. Income inequality, on the other hand, reduces private contributions by a significant margin. In the second chapter, my coauthors and I examine how job transfer rules and preferences affect labor market efficiency and access to quality teachers. To do so, we recover teacher and school preferences using data from Minneapolis Public Schools' web-based internal teacher labor market. Overall, we find that the average teacher prefers schools serving already-advantaged students and the average school prefers applicants who are more effective, hold an advanced degree, and not in their early-career. These preferences help explain why we observe the troubling sorting patterns among teachers and suggest that further liberalizing the teacher labor market may exacerbate the inequitable distribution of quality teachers. Finally, the third chapter evaluates a hidden social cost of air pollution beyond hospital admissions and premature deaths: student achievement. Given the strength of evidence linking academic performance to long-term life outcomes and the fact that disadvantaged and marginalized communities tend to get more exposure to air pollution, this additional cost should be identified and quantified correctly. Using an exogenous source of variation in the levels of air pollution from the closure of an airport terminal, I find that the closure led to a roughly 2 percent of a standard deviation increase in high-stakes test scores.

Table of Contents

List of Tables

List of Figures

Chapter 1:

Heterogeneity and the Private Contribution of Local Public Goods: Evidence from NYC311

I. Introduction:

The United States is on the verge of a diversity explosion. The United States Census Bureau projected that the US will become a country with no racial majority by 2044 (Frey 2014). The trajectory of the level of heterogeneity in the income dimension is equally striking. From 1980 to 2014, the top 1 percent adults went from accounting for 12 percent of national income to over 20 percent while the bottom 50 percent adults saw their national income share drop from 20 percent to 12 percent (Piketty et al. 2016). New York City (NYC), America's densest city and historically an important port of entry for immigrants, has been at the forefront of this diversity explosion. Roughly 38 percent of NYC residents are now foreign-born (United States Census Bureau 2015). On average, there is a fifty percent chance that two randomly selected NYC residents will belong to different racial groups and speak different languages at home. The Gini coefficient in NYC is also among the highest (United States Census Bureau 2015). In Manhattan, income inequality is comparable to the levels in Brazil and Colombia.

Social scientists have long debated whether heterogeneity is good for society. The literature remains inconclusive at best as to whether there is a net societal benefit from having a high level of urban heterogeneity. Some empirical studies support the Jacobs (1969) hypothesis that urban heterogeneity offers a more favorable environment for economic development (Ejermo 2005; Glaeser et al. 1992). On the other hand, Putnam

(2007) points to diversity as the main culprit for reducing social capital in US cities. In his view, too much diversity can erode trust and may exacerbate collective action problems. Alesina, Baqir, and Easterly (1999)'s finding that ethnic fragmentation hampers the provision of public goods in the US started an entirely new empirical literature that would eventually discover other negative effects of ethnic heterogeneity such as conflicts (Reynal-Querol and Montalvo 2005) and economic underdevelopment (Alesina et al. 2003). Studies in experimental economics find that even the *private* provision (or private contribution) of public goods can be negatively affected by heterogeneity (Chan et al. 1999). While there is no consensus on the net societal benefit of heterogeneity, recent findings indicate that there is a negative relationship between heterogeneity and both the public and private provision of public goods.

Against the backdrop of high level of heterogeneity found in modern cities and the prevailing view of its effect on public goods provision, testing for and understanding the relationship between heterogeneity and the private contribution of local public goods (hereinafter referred to as "private contributions") in a context outside of a laboratory setting is the objective of this paper. In particular, I estimate the effects of heterogeneity on citizens' propensity to make non-emergency service requests in NYC that benefit not only the contactor but also his or her neighbors. Each year between 2005 and 2014, the NYC311 call center handled over 1 million requests and complaints about the quality of local public goods ranging from traffic signal problems to noise complaints. Citizens can make requests to NYC311 by calling the number 311, reporting online, or using the NYC311 mobile applications. The various departments within the city then rely on this database that stores information about the type of request, exact location, and time of request to locate and restore the quality of the reported local public goods in a timely manner. Thus, citizen participation in NYC311 translates indirectly to the production or the restoration of local public goods. I view this action as an economic agent voluntarily contributing to a local public good.

To estimate the impact of heterogeneity on private contributions, I construct a census tract by month panel data set from 2005 to 2014 containing information on the frequency and type of service requests, the various measures of heterogeneity, and the socioeconomic characteristics of NYC residents. For my dependent variables, I carefully select the types of requests, geocode, and aggregate more than 2 million service requests to the month level by request type for each census tract. Then, I link these data with the heterogeneity measures and other control variables constructed from two rounds of the American Community Survey (ACS) 5-year estimates: 2005-2009 and 2010-2014.

Identifying the effect of heterogeneity on the propensity to make service requests is challenging since not all problems with local public goods are observed and they are not always random across geographical areas and time. Some areas may have higher needs and some areas may receive better treatment from the city. To address these estimation issues, I employ three empirical strategies. First, I focus my analysis only on the types of requests whose causes are exogenous given my econometric specification and which are unlikely to be requested by nonresidents. Second, I exploit the panel nature of the data set and employ a two-period fixed effects model using both the census tract-specific and time fixed effects. I also include a rich set of time-varying control variables to address other potential confounders such as tastes for public goods, mobility, and service completion time. With the first two strategies, the level of need to contact NYC311 across geographical areas and time should be adequately controlled for and the occurrence rate at which of these public goods incidents happen should be as good as random. Third, I instrument income inequality to address concerns about measurement errors and reverse causality. In survey data like the ACS, measurement errors can be large, and fixed effects models with slow-moving variables can exacerbate attenuation bias. Moreover, it could be that it is the change in the quality of local public goods that leads to the change in income inequality, not the other way around. Richer residents may choose to relocate to areas with better public goods (areas requiring few 311 requests), mechanically changing the level of income inequality both at the origin and the destination. Three instruments are used to address these concerns. They are the fraction of households in the top income bracket, the fraction of workers in the information industry, and the fraction of workers in the public sector for each census tract in 2000, several years before the two rounds of ACS. The intuition is that the income of top earners and workers in the information industry grew disproportionately faster over time, while the income of workers in the public sector should not be as affected by the national patterns of income growth or by the Great Recession as was the income of workers in private industries. By freezing these fractions in 2000, I foreclose the possibility that the quality of local public goods causes the change in income inequality by attracting new residents. Given a rich set of controls, these instruments should not affect service requests except through income inequality.

Overall, I find little to no evidence that racial and linguistic heterogeneity reduce private contributions. While the estimates are mostly negative across multiple specifications, they are generally neither statistically different from zero nor economically significant. One reason is that they are highly correlated with income inequality and other socio-demographic variables. In all of my specifications, income inequality is negatively associated with private contributions. My preferred IV estimate suggests that a 0.01 point increase in the Gini coefficient—roughly the same magnitude that the average Gini coefficient changed over a five-year period in NYC— in an area reduces private contributions by roughly 2.5 percent. This effect is robust to a number of alternative specifications, including using a different set of request types as the dependent variable and running the model over each round of the ACS separately. Median household income and population are negatively associated with private contributions but their effects tend to be economically insignificant.

The contribution of this paper is threefold. First, this paper is the first to use the 311 service request data to study the private contribution of local public goods. Unlike recent studies that focus on the geographical distribution of government services (Levine and Gershenson 2014; Feigenbaum and Hall 2015), on predicting public goods problems (Hsieh, Yen, and Li 2015), or on the determinants of neighborhood conflicts (Legewie and Schaeffer 2016), I present a set of empirical strategies that allow researchers to treat service requests as an act of private contribution of local public goods. This opens a wide range of possibilities for future research in the areas of public economics as more cities adopt the 311 system and open their data sets to researchers. Particularly, the 311 context offers several advantages over other types of observational data. Because there is no explicit economic incentive for citizens to make service requests *other than* the utility from the problem being resolved as a result of the call and the disutility of effort, the observed voluntary contribution is independent of other potential monetary factors that play a key role in studies that treat charitable donations, gifts, and political contributions as measures of private contributions to public goods (Andreoni and Payne 2013).

Second, this paper adds new insight and depth to the literature on the relationship between heterogeneity and public goods provision. One reason why my results run counter to the prevailing view in the literature is that the local public goods selected in this paper do not have the redistributive properties and they have little to do with the differing tastes among heterogeneous populations. In other words, racial and linguistic heterogeneity does not reduce private contributions in this context where taste and redistribution are unlikely to be important. Rather, it is income inequality that lowers private contributions. This result illustrates the danger of over-generalizing the prevailing theme in the literature to all types of public goods.

Finally, this paper makes a contribution to the ongoing debate about the underlying mechanism behind the relationship between heterogeneity and private contribution. Currently, the mechanisms are not well understood and often are impossible to pinpoint outside controlled public goods games (Habyarimana et al. 2007). The specific nature of how service requests are made through the NYC311 system presents a rare opportunity to identify the underlying mechanism. Because making service requests is free and the local public goods studied in this paper do not have redistributive properties, most traditional explanations of why income inequality would affect private contributions are ruled out. Due to the nature of how service requests are made and handled, there is also little room for social norms such as sanctioning to influence private contributions as would be in other contexts. Therefore, I argue that my results are primarily driven by the taste of the beneficiary explanation where agents yield less utility when out-group agents consume the local public goods.

The rest of the paper proceeds as follows: Section II discusses the conceptual framework and relevant literature; Section III presents the NYC311 context and the data; Section IV discusses the rest of the data set; Section V outlines the empirical strategies; Section VI presents the empirical results; Section VII concludes the paper.

II. Conceptual Framework

Although the negative relationship between heterogeneity and public goods provision is well known, the underlying mechanisms are not well understood. Habyarimana et al. (2007) identify at least three families of mechanisms: preferences, technology, and strategy. Within the preferences family, there are two submechanisms. The first submechanism posits that the lack of commonality of preferences reduces the level of public goods in relatively more heterogeneous communities. Heterogeneity, in this sense, implies diversity of taste. In heterogeneous communities, agents prefer different goods and services, so they prefer to pull fewer resources together to fund public projects. Alesina, Baqir, and Easterly (1999), perhaps the most well-cited study in this area, develops a model that captures this behavior and illustrates this mechanism empirically with data on shares of public goods spending from US cities. Unlike the commonality of taste explanation, the second submechanism does not put any restriction on agent preferences regarding the nature of the public goods, but on agent preferences regarding the beneficiary. Often called other-regarding preferences or taste for the beneficiary, this submechanism involves agents who attach relatively less utility (or even negative utility) when agents from outside of their group consume the public good. As Becker (1957) puts it, this is clearly a "taste for discrimination." In the NYC311 context, if there is any relationship, it is more likely that the taste for the beneficiary mechanism is at play since the types of local public goods selected for this study are basic necessities for daily urban life. Contextual examples from Alesina, Baqir, and Easterly (1999) and Miguel (2001) involve different ethnic groups demanding different types of public goods that would benefit themselves more than out-group members. These include public education spending in areas with high income inequality and diverse ages, and religious holidays and language of instruction at school in culturally diverse communities. In contrast, most service requests are about everyday necessities.

The second family of mechanisms, technology, concerns the efficacy of collective action in a community. Deutsch (1966) and Hardin (1995) argue that homogeneous communities may be more successful at public goods provision since it is easier for co-ethnics, due to their common language and culture (having a better "technology"), to communicate among themselves and reach a consensus. In the NYC311 context, I argue that this is unlikely to be the underlying mechanism. Although the collective action problem remains, there is no difference in the technology of how agents in communities with different levels of heterogeneity decide whether to make service requests or ignore the problems. First, for a problem to be resolved, contacting NYC311 does not require citizens to coordinate or vote to reach a consensus. Everyone makes a decision voluntarily, individually, and privately. Second, there is no obvious set of strategies to contact NYC311 that is available to in-group members but not to out-group members. NYC311 is available to all residents in the New York City metropolitan area regardless of their race, ethnicity, or spoken language. There are translators to handle over 180 different languages. Therefore, there should be no difference in the level of the efficacy of collective action among different census tracts.

The third family of mechanisms is strategy-based. Strategy-based explanations focus on how beliefs influence agents to form a strategy to contribute or free-ride. For example, in a simple game with two choices, contribute or not contribute, if all players believe that in-group members will always contribute and out-group members will always not, the labeling of which groups the players belong to can influence the selection of strategies and outcome of the game. Habyarimana et al. (2007) champions the most prominent submechanism in the political science literature: social sanctioning. It involves two essential elements: a norm of cooperation within a group, but not across groups, and a credible threat of sanctioning if any in-group member shirks. In the NYC311 context, however, it is difficult to imagine that there can be any credible threat of sanctioning as seen in the Kenyan context of Miguel and Gugerty (2004). This aspect of strategy-based mechanisms is absent. The reasoning is similar to why technology-based mechanisms are unlikely. Contacting NYC311 is a private act and so the players cannot possibly punish a specific person for not making service requests. Therefore, if there is any negative relationship between heterogeneity and private contribution, it likely will stem from the taste for the beneficiary mechanism.

III. The NYC311 Context and Service Requests Data

The NYC311 system is open all day, every day, and is the largest social services information center in the world (Wiseman 2015). It receives more than 3,600 kinds of service requests, linking together every department associated with each service so that citizens just need to remember one number regardless of whose jurisdiction each incident falls under. Since its inception, all non-emergency service requests have to go through NYC311, not through the individual departmental systems. Citizens can self-report the type and location of the incident or they can also just take a picture and submit it through an automated system. The NYC311 system currently handles over 150 languages and integrates this user-input information with agency work order management systems to address service requests.

I obtained the incident-level 311 data for NYC from OpenData.gov. While there are thousands of types of requests, I focus only on two types of requests: sewer blockage and street light outage.[1] I select this set of request types to minimize the likelihood that their rates of occurrence can be affected by unobserved factors beyond what my specifications can capture. Given census tract specific fixed effects or a set of proxy

[1] Most types of local public goods are financed almost exclusively by revenue from property and other taxes at the city level (NYC IBO, 2013). In other words, the costs of restoring the public goods are already sunk and thus should not affect the individual decision whether to contact NYC311.

variables such as the number of households and population, these types of requests occur in a relatively more exogenous fashion than others such as requests about noise or potholes. While sewer blockage and street light outage are typically caused by severe weather conditions, requests such as potholes occur as a function of both the weather and other known, but unobserved, factors such as vehicle-miles driven and vehicle weight per year. For robustness checks, I also include an alternative set of requests about power outage and heating complaints.

I then map latitude and longitude (and in some cases street intersections) of every single request of the four types from 2005 to 2014 to the census tract it belongs to using TIGER/LINE shape files and ArcGIS. Simple geospatial analysis suggests that there is a large degree of geographical variation in both the frequency of requests and the level of heterogeneity. Figure 1.1 depicts the relationship between the Gini coefficient and the frequency of service requests per capita from 2010-2014. Areas that have high level of income inequality tend to have low frequency of service requests. In contrast, there seems to be no such negative relationship for racial and linguistic heterogeneity (Figures 1.2 and 1.3).

There is some variation in the frequency of monthly service requests by time. Table 1.1 shows that the per capita frequency of requests declined over the two ACS rounds and the differences are statistically significant. It is possible that since the beginning of the NYC311 system, problems are less likely to be generated overtime because many of them have already been resolved. Another possibility is that citizens make fewer requests because they perceive them to have little impact. For both of these, as well as other, possibilities, a time trend must be controlled for.

IV. Data on Heterogeneity and Socioeconomic Characteristics

I obtain data on heterogeneity and other socioeconomic characteristics from two rounds of ACS 5-year estimates (2005-2009 and 2010-2014). For racial and linguistic heterogeneity, I use the reverse Herfindahl–Hirschman Index (HHI) to construct the measures of fractionalization. The HHI is widely used in the literature as a measure of ethnolinguistic fractionalization (ELF) due to the appeal that it can simply be interpreted as the probability that any two randomly selected individuals in a particular area belong to different racial or ethnic groups (Montalvo and Reynal-Querol 2010). In other words, it captures the likelihood that a person associated with a particular group will encounter someone who is an "outsider." Inverse HHIs are calculated with the following formula: $1 - HHI = 1 - \sum_i {s_i}^2$. For racial fractionalization, s_i is the fraction of each race recorded in the ACS. For linguistic fractionalization, s_i is the fraction of residents who speak one of the five types of languages recorded in the ACS. The higher is the HHI, the higher is the degree of fractionalization. In the income dimension, I refrain from using the HHI in favor of the Gini Coefficient (Gini) because the HHI does not capture the depth of income inequality as well as the Gini.[2]

[2] While the HHI is an appropriate measure for racial and linguistic heterogeneity, it can be too coarse for income inequality. The ACS breaks households into nine income brackets and the top bracket is censored at over $150,000. The HHI can severely underestimate the true level of income inequality in census tracts that have many top income households or a few households ultra-high earners. Using the HHI to measure income inequality may also waste useful information about the distance between households with different

According to Table 1.2, these heterogeneity measures, on average, all increase over time, confirming prior beliefs that there are both geographical and temporal variations among the census tracts. NYC is heterogeneous in all of these dimensions. There is almost a 50 percent chance that two randomly selected people will belong to different racial groups and will speak different languages at home. The rise in income inequality over this period is often attributed to the disproportionate effects of the Great Recession and the recovery on different income groups (Saez 2015). Table 1.3 shows the sample mean and standard deviation for key control variables in this study over the two sampling periods. The most prominent changes over time are that the fraction of white population declined while the unemployment rate and the fraction of adults with at least a Bachelor Degree rose considerably.

V. Empirical Strategies

Estimating the causal impact of heterogeneity on private contributions using observational data is challenging due to endogeneity, measurement errors, and reverse causality. To address these challenges, I employ three empirical strategies: 1) restricting the types of service requests; 2) including census-tract fixed effects and a rich set of time-varying control variables; and 3) using instrumental variables for income inequality.

To alleviate concerns about endogeneity, I first carefully select the types of service requests. As outlined in Minkoff (2016), the frequency of service requests is driven by two

incomes. The distance between each race or each spoken language has little objective meaning, but the distance between two monetary values does have interpretable meaning that should be used in the analysis.

distinct independent forces: the spatial conditions within a given area and the citizens' propensity to contact NYC311. Without adequately controlling for the former, which essentially is a risk set, estimates of the latter would be biased. For example, a densely populated census tract may have more occurrences of certain types of incidents such as damaged traffic signs just because it has more intersections than less densely populated census tracts. Likewise, a bad storm year can lead to a greater number of service requests about clogged drainages or short circuited traffic signals in all census tracts regardless of the variation in heterogeneity measures. Additionally, problems such as traffic signal malfunction, although occurring randomly by nature within a given area, risk being reported by travelers, rather than the actual residents whose characteristics I observe from the ACS. Therefore, I restrict my analysis to two sets of public goods problems to avoid these concerns. The main set of requests for the analysis includes requests about sewer blockage, sewer backup, and street light outages. These problems typically occur because of severe weather conditions such as heavy rain, melted snow, or a storm. I also construct an alternative set of requests that include requests about power outage and heating complaints. Compared to the main set of requests, these requests are arguably more private in nature, but the act of monitoring and requesting government service still captures the essence of the definition of the private contribution of local public goods because these problems often affect multiple residential homes and apartments simultaneously.

To further reduce concerns about omitted variable bias, I employ census-tract specific fixed effects. These fixed effects would compare each census tract to itself over time and difference out all time-invariant unobserved neighborhood characteristics such as the infrastructure of the streets, the number of intersections, and the structure of pipe networks as well as city's preferential treatment towards certain areas. I also introduce flexibility into the model by including the year-month fixed effects to control for aggregate shocks, seasonality, and the level of familiarity with the NYC311 that may affect every census tract in NYC over time.

Formally, the fixed effects specification can be expressed as:

$$Y_{it} = P_{it}'\beta + X'_{it}\gamma + Tract_i + my_t + \epsilon_{it} \quad (1)$$

where the dependent variable, Y_{it}, is the frequency of per capita monthly service requests about local public goods in census tract *i*, in month-year *t*; P_{it} is a vector of heterogeneity measures; X_{it} is a vector of the covariates described in Table 3 and the fractions of workers in every industry; $Tract_i$ is the census tract fixed effects; my_t is the month-year fixed effects; and ϵ_{it} is the random disturbance term.[3] Because unobserved time-varying citizen characteristics can cause bias if they codetermine both heterogeneity and the number of service requests, I include a rich set of socioeconomic characteristics controls in the vector X_{it}. Median household income, total population, percent of adult population with a Bachelor's degree of higher, travel time, age composition, racial composition, industry composition, spoken language composition, percent of population that is unemployed, percent of household married with at least one child, percent of population that did not

[3] Note that the census tract-level variables from the ACS vary by round but not by month-year as does the dependent variable. In summary, the dependent variables vary by month-year for each census tract from January 2005 to December 2014, but the explanatory variables for each census tract remain the same from January 2005 to December 2009, vary once, and remain the same from January 2010 to December 2014.

recently relocate, percent of foreign born population, and percent of owner occupied homes are examples of the factors that may determine citizens' propensity to contact 311 (Minkoff 2016). Equation (1) is estimated using Ordinary Least Square (OLS).[4]

Another potential challenge to the identification is that local governments may treat certain census tracts better than others. The heatmap in Figure 1.4 should put aside much of any concerns about unequal treatment among census tracts. There is no strong correlation among completion time and socioeconomic characteristics of census tracts such as income and racial composition. This is not a surprise since the goal of NYC311 was to standardize the quality of government services, and census tracts are not political boundaries. Still, I include the completion time variable in my set of control variables to absorb any other impact it may have on the propensity to use NYC311. For example, if it takes a long time to resolve local public goods problems in a particular census tract, residents may not use NYC311 as much just because they feel that the system is ineffective, not because of the level of heterogeneity changes in their census tract.

Although the fixed effects model can help address many concerns about endogeneity, it does not address the issues of measurement errors and reverse causality. Both issues are likely to be important given the nature of survey data and the context of local public goods. More importantly, income inequality, unlike racial and linguistic

---

[4] I choose to estimate all of my specifications using OLS with fixed effects rather than with count models for the following reason. First, count models suffer from incidental parameter bias. Second, the negative binomial with fixed effects estimates are not true fixed effects estimates (Hausman et al. 1984). As discussed earlier, having a reliable empirical strategy such as the fixed effects model is crucial for drawing valid statistical inferences in this paper. To address the excessive zero problem with count data, I follow Burbidge (1988) and use the Inverse Hyperbolic Sine transformation (IHS) to transform the monthly frequency of service requests. With IHS, unlike taking the logarithm, zeros are defined.

fractionalization, is hardly exogenous. If the Tiebout hypothesis is correct, an increase in income inequality may just be the result of households sorting across neighborhoods for an optimal bundle of local public goods according to their tastes. Specifications that do not address this possibility can suffer from reverse causality.

Fortunately, the empirical literature suggests that sorting between neighborhoods in US cities has remained fairly constant or even declined in recent years (Cutler, Glaeser, and Vigdor 1999 and Kremer 1997). In the context of this paper, the bias from Tiebout sorting is also likely to be less of a concern due to three reasons. First, the local public goods studied in this paper are standard basic public amenities that are maintained at the city level. In contrast to public education that is provided with a much larger degree of variation in quality, it is unlikely that these local public goods alone can drive relocation trends. Second, a measure of mobility is already controlled for in my set of control variables. It also varies little over time, confirming the findings from the neighborhood sorting literature.[5] Third, the quality of government responses to these requests is likely to be similar across census tracts since my analysis of completion time for service requests reveals no preferential treatment among census tracts.

While recent empirical findings and the context of this paper provide some reassurance, they cannot fully rule out the Tiebout hypothesis. Most specifically, the possibility that citizens may relocate to consume other local public goods that are correlated

[5] Even with little mobility between the two rounds of ACS, the variation in heterogeneity can still come from other ways. Income inequality can change over time because household income may grow at different rates. In the racial and linguistic dimensions, the level of heterogeneity can vary as population grows or shrinks.

with the unobserved quality of the local public goods studied in this paper cannot be completely ruled out. For this reason and to correct for attenuation bias, I instrument the Gini coefficient for the distribution of income observed in 2005-2009 and 2010-2014 with a group of instruments from the 2000 Decennial Census.

The first instrument is the fraction of households earning more than $200,000 in 2000. Studies of national income growth trend in the US suggest that income at the top of the income distribution grew disproportionately faster than did income for the rest of the society (Saez 2015). I expect census tracts with a larger share of top income households to get more exposure to this trend and thus experience a higher level of income inequality over time. The second instrument is the fraction of workers in the information technology industry in 2000. I exploit one of the most common explanations for income inequality in the post-war era: skill-biased technological change (SBTC). The shift in demand for high-skill labor and the rapid growth in information technology during this period favored workers in this industry (David et al. 2006; Heathcote et al. 2010). Riding the wave of technological advancements, census tracts with a large fraction of workers in the information industry in the past should experience more income inequality over time. Unlike the first two instruments, the third instrument rests on the flip side of this intuition. I expect census tracts with a larger share of workers in the public sector to experience a decline in income inequality over time since the within-tract workforce composition should get a relatively little exposure to both the national income growth trend and the SBTC. The income distribution in census tracts with a larger share of public sector employees should also be more immune to the impact of the Great Recession and its uneven recovery that spurred the rise in income inequality in recent years (Kopelman and Rosen 2016; Saez 2015).

The identifying assumption is that these instruments do not affect the frequency of service requests per capita except through the Gini coefficient, given other control variables in my model. Since all of the instruments are from 2000, I mechanically preclude the possibility of reverse causality. Controlling for multiple socio-demographic variables such as current income, month-year fixed effects, and the fractions of workers in every industry that the ACS collects, I argue that the instruments do not affect other unobserved determinants of service requests that may still remain in the error term. In the results section, the Sargan-Hansen overidentification test will provide another screening test whether the excluded instruments are appropriately independent of the error process.

Figure 1.5 shows the relationships between the instruments and the frequency of service requests per capita. The fraction of top income households and the fraction of workers in the information industry in 2000 are negatively correlated with service requests, while the fraction of workers in the public sector in 2000 is positively correlated with service requests. These relationships are not surprising given the ways I expect the instruments to affect the current level of income inequality. To visualize the strength of each instrument, I plot them against the Gini coefficient observed in the two rounds of ACS in Figure 1.6. As expected, the fraction of top income households and the fraction of workers in the information industry in 2000 are positively correlated with the Gini, while the fraction of workers in the public sector in 2000 is negatively correlated with the Gini.

The 2SLS equations are given below:

$$Gini_{it} = Z'_{i2000}\theta + H'_{it}\beta + X'{}_{it}\gamma + my_t + \eta_{it} \quad (2)$$

$$Y_{it} = \alpha\widehat{Gini}_{it} + H'_{it}\beta + X'{}_{it}\gamma + my_t + \mu_{it} \quad (3)$$

where the dependent variable in the first-stage regression, $Gini_{it}$, is the Gini coefficient in census tract i, in month-year t; Z_{i2000} is a vector of the instruments taken from the 2000 Decennial Census; H_{it} is a vector of racial and linguistic heterogeneity measures; X_{it} is the vector of covariates described in Table 3; my_t is the month-year fixed effects; η *and* μ is the random disturbance terms; and $\widehat{Gini}$ is the predicted Gini coefficient from the first-stage.

VI. Results and Discussion

The theoretical discussions in Section II suggest that the relationship between heterogeneity and private contributions in the NYC311 context might be a negative one. Across multiple specifications, I find robust empirical evidence that it is income inequality that lowers private contributions, not racial or linguistic heterogeneity. Total population and median income are both negatively associated with private contributions in most specifications.

Table 1.4 presents the results from estimating variants of equation (5). In a basic OLS specification with no controls or fixed effects, the racial fractionalization index, along with the Gini coefficient, have statistically significant negative coefficients. However, once the month-year fixed effects and controls are included in column 2, the coefficient on the racial fractionalization index becomes much smaller and is no longer statistically different from zero. The magnitude of the coefficient of the Gini coefficient also drops by half, suggesting that both the time shocks and omitted variable bias may be at play, but it remains highly statistically significant. Column 3 introduces the census-tract specific fixed effects with no controls, and the magnitude of the effect of income inequality drops further towards zero, though it is still statistically significant at the 5 percent level. Adding control variables in column 4 does not significantly change this magnitude. The drop in the magnitude of the Gini coefficient from close to -1 in column 2 to around -0.45 in columns 3 and 4 could be from either controlling for time-invariant unobserved characteristics or from a worsened attenuation bias. Since fixed effects models tend to exacerbate attenuation bias, actual impacts may be substantially larger in magnitude. With all the fixed effects and control variables included, a 0.01 point increase in the Gini coefficient is associated with roughly a 0.5 percent reduction in private contributions. Back-of-the-envelope calculations suggest that going from the level of income inequality in Manhattan, where income inequality is the highest, to that of Queens would lead to a roughly 10 percent increase in private contributions.

Across multiple specifications, the negative effect of total population is not surprising. Residents can free-ride others to make service requests for them. A one percent increase in the population is associated with a decline in private contributions from 0.4 to 0.7 percent. This can be interpreted as a real-world evidence of the bystander effect first coined by Latane and Darley (1968). A more surprising finding is that median income is negatively associated with private contributions even though the level of need for service should be as good as random across census tracts. This contradicts the findings from Feigenbaum and Hall (2015), who find that high income neighborhoods are more likely to use NYC311. The presence of this income effect may imply that residents in higher income areas tend to also have higher opportunity costs.

In Table 1.5, results from 2SLS specifications are presented. All specifications include the month-year fixed effects and controls. In columns 1 and 2, I separately instrument the Gini coefficient with the fraction of top income households and the fraction of workers in the public sector in 2000, respectively.[6] For both specifications, the first-stage F-statistics suggest that the instruments are strongly correlated with the Gini coefficient. The strength of the fraction of top income households as an instrumental variable is especially notable since the first-stage F-statistics is over an order of magnitude larger than the conventional value of 10 for a "strong" instrument. The estimates of the impact of income inequality rose considerably from OLS estimates, to -2.3 and -4.9 in columns 1 and 2, respectively. In columns 3 through 5, I use both instruments at once to potentially get more asymptotic efficiency. The model in column 3 is run over the entire sample while the models in columns 4 and 5 are run separately for each round of the ACS. The Sargan-Hansen test results all indicate that the null hypothesis that the instrumental variables are not correlated with the error term in these models cannot be rejected. The estimates in column 3 are my preferred estimates due to a large first-stage F-statistics and the sample size. With 2SLS, a 0.01 point increase in the Gini coefficient—a change

[6] The fraction of workers in the information industry does not pass the conventional F-statistics test. Therefore, I exclude it in this specification. However, it will be used in a robustness check specification with an alternative set of requests as the dependent variable.

comparable to the actual increase of the average Gini coefficient between the two rounds of ACS in NYC— leads to roughly a 2.5 percent reduction in private contributions.

As an additional robustness check, I run both the fixed effects model and the 2SLS model with an alternative set of service requests which include requests about electrical outage and heating complaints. In Table 1.6, the estimate of the effect of the Gini coefficient is not statistically different from zero in the fixed effects model, but the 2SLS estimates are comparable to my preferred 2SLS estimates in Table 5 once I instrument the Gini coefficient with the fraction of top income households and the fraction of workers in the information industry in 2000. Generally, the magnitude of the 2SLS estimates with an alternative set of service requests is slightly smaller.

My findings run counter to the classic theoretical prediction that the private provision of public goods is independent of income redistribution (Bergstrom, Blume, and Varian 1986; Warr 1983). The reason is that the public goods studied in this paper are not pure public goods. They are locally consumed, require very little effort to contribute, and can be congested.[7] My results are more in line with the recent findings from studies such as Anderson et al. (2008) and Tavoni et al. (2011). In these studies, income is exogenously varied among players in a laboratory seting and the authors find that income inequality reduces the private provision of public goods because it erodes social capital. While the contexts of these studies differ greatly, they both point to a reduction in social capital as a common underlying mechanism. The negative relationship between income inequality and

[7] Allouch (2015) provides a formal proof that many neutrality results in the classic theoretical works on the private provision of public goods do not hold when public goods are no longer pure.

social capital is also found in a large scale observational study by Alesina and La Ferrara (2000) where social capital is proxied for by the various membership participation rates at the metropolitan level. In Canada, Payne and Smith (2015) finds that income inequality reduces charitable giving, which is also a form of privately provided public goods.

Since the local public goods in this paper do not have redistributive properties and cost very little effort, my results—taken together with the insights from these studies—suggest that income inequality likely reduces private contributions through its effects on the utility individuals attach to the welfare of their neighbors. Given the consistency of my estimates, reducing the level of income inequality can be an effective policy option to encourage more private contributions along with the many improvements in social outcomes that have already been extensively studied such as social cohesion and public health (Pickett and Wilkinson 2015). More importantly, the results in this paper suggest that previous findings on the societal costs of racial and linguistic heterogeneity do not necessarily generalize to other settings.

VII. Conclusions

This paper estimates the effects of heterogeneity on private contributions to local public goods. It takes a series of steps to provide rigorous empirical evidence on the relationship between heterogeneity and private contributions. Combining the NYC311 data with two rounds of the ACS, I employ census-tract fixed effects models and cross-sectional 2SLS with three instrumental variables. I find evidence across multiple specifications that it is income inequality, not racial or linguistic heterogeneity, that lowers private contributions to local public goods. As communities become more heterogeneous in the income dimension, individuals contribute less, on a per capita basis, to local public goods. In my preferred 2SLS specification, a one standard deviation increase in the Gini coefficient reduces private contributions by about 33 percent. Back-of-the-envelope calculations suggest that raising the level of income inequality in Queens to the level in Manhattan (a change of about 1.5 standard deviation) would bring about a 49 percent reduction in private contributions. Because of the specific nature of how service requests are made through NYC311, the underlying mechanism is likely the taste for the beneficiary mechanism in which individuals care less about their neighbors' welfare.

There are two important caveats to keep in mind when interpreting and extrapolating these results. First, due to the many steps taken to improve the statistical inference in this paper, the interpretations are quite specific to the NYC311 context and the types of public goods chosen for this paper. Had different types of public goods been chosen, or the mechanics of how residents make contributions were different, the underlying mechanism and the overall effect on private contributions may have been different. Second, the fact that some of my results vary when I use an alternative set of the types of requests suggests that more research is needed to understand how and why the specific nature of each type of public goods can give rise to these differences. Excludability of these local public goods is likely an important determinant of these differences. Finally, pairing service requests data with subjective wellbeing data or home prices data may well allow researchers to estimate willingness to pay measures for many types of public goods beyond those for which the demand has already been well documented.

Chapter 2

School and Teacher Preferences: Evidence from a Multi-stage Internal Labor Market

(with Elton Mykerezi and Aaron Sojourner)

I. Introduction

Is a completely free teacher labor market really a good idea? As more US school districts begin to give schools more hiring autonomy and reduce the importance of seniority, the distribution of teacher quality remains a major concern among policymakers. High-performing, high-experience teachers tend to be concentrated in high-achievement schools mostly serving students from relatively advantaged backgrounds (Isenberg et al. 2013). From the perspective of equity, this situation is troubling since quality teachers are not where they are needed the most. Previous studies have pointed to teacher mobility as one of the primary causes of this phenomenon. Because salaries are mostly flat within a school district, schools with less desirable nonmonetary characteristics tend to have a difficult time competing for, and retaining, high quality teachers. Moreover, transfer rules that favor seniority can reinforce this trend, allowing schools that are generally perceived to be more desirable to recruit even more experienced teachers from other schools. Although some studies have used work history data to identify these transfer patterns in many school districts, observing only the final matched outcomes between teachers and schools do not allow for understanding why, and at what point in the hiring process, these patterns emerge. More importantly, data limitations often prevent researchers from identifying the extent to which preferences and other labor market rules contribute to these patterns. Without a good understanding of how these factors shape the final matched outcomes, it is difficult to design any policy to address concerns about disadvantaged students' access to quality teachers.

Data on the entire job transfer process allows one to gain more insights into the inner workings of the teacher labor market. This paper analyzes the two-sided multi-stage matching process between teachers and schools in the Minneapolis Public Schools (MPS) that occurs on a web-based internal teacher transfer platform where every action from every market participant is recorded. On the teacher side, these actions include which teachers apply to which postings from which schools and which job offers they accept or decline. On the school side, we observe how they select candidates for interviews at the resumé-screening stage and how they rank candidates at the end of the interview stage. Linking these actions with teacher and school characteristics information recorded in the administrative data set allows us to separately recover teacher and school preferences.

Our contribution is three-fold. First, we add much-needed evidence to a handful of studies that can separately identify teacher and school preferences (Boyd el al. 2011, 2013). Across multiple specifications of our discrete choice models, we find that an average teacher prefers schools with high proportions of white students and teachers, high reading achievement, high average teacher experience, high enrollment, and low pupil-teacher ratios. All else equal, teachers prefer less commute time. We also find evidence of heterogeneity in teacher preferences. Although on average teachers do not find schools with high proportions of students of color desirable, teachers of color prefer them more than white teachers do. Highly-effective teachers, on the other hand, dislike schools serving high proportions of students of color than less-effective teachers. For school preferences, we find that being white, female, young, more effective, and holding an advanced degree increases the likelihood of being hired. Schools across the spectrum of student advantage hire applicants who are substantially more effective than their average applicant, despite the fact that they do not systematically know applicants' effectiveness ratings.[8] This is evidence that all schools can recognize and do value effectiveness among applicants.

Second, this paper is the first to estimate the changes of hard-to-staff school characteristics required to attract more highly effective applicants in a revealed preference framework. Using the estimated teacher preferences, we compute the marginal rates of substitution among school characteristics and the changes in the school characteristics required to create indifference between working at any type of schools and for any type of teachers. To demonstrate a possible policy change, we estimate that the district must increase the pay at bottom-quintile schools by roughly $1,500 a year to create indifference for highly effective teachers to transfer from top-quintile schools. This increase in pay scales up considerably as commute time increases.

Finally, our paper takes advantage of the multi-stage nature of the job transfer process at MPS and finds evidence that school preferences before and after the interview stage are different. At the resumé-screening stage, schools highly prefer candidates whose resumés indicate that they are young, white, holding an advanced degree, experienced, effective, and have no history of school-hopping. As schools obtain more information in the interview stage, the importance of advanced degree, experience, and history of school-

[8] Our interviews with five school principals confirm that schools do not know applicants' effectiveness. In addition, in all years of I&S not one applicant shared her effectiveness ratings at any time with these principals.

hopping diminishes while age, being female, being an early contract hire, and effectiveness show greater importance.

This chapter is organized as follows. Section II provides the background and context of the educational landscape in MPS and its online job transfer platform. Section III gives an overview of the findings in the related literature. Section IV lays out the conceptual framework that provides a foundation for the empirical strategies, which are discussed in detail in Section V. Section VI presents the results, after which Section VII concludes.

II. Background and Context

Interview & Select (I&S) is a web-based internal teacher transfer platform for MPS system, a school district that covers all of the city of Minneapolis, Minnesota. The total enrollment in MPS is around 35,000 students in primary and secondary schools. It is among the most diverse school districts in the US with students speaking over 90 different languages at home. Like in many other US school districts, schools serving high fractions of students of color tend to have low student achievement and they employ teachers who are less experienced and less effective (Figure 2.1).[9]

At MPS, vacancies are filled internally first. After the internal job transfer process is over, the district then places external applicants to any remaining position not taken by

[9] Throughout this paper, we use the growth in the percentage of students who are proficient in reading that partials out any prior student achievement as measure of achievement in a given school year.

its existing teachers. As such, teacher and school preferences in the internal labor market are more likely to have an effect on the distribution of teachers. Instead of giving the most senior teacher a guaranteed placement at her desired school as in the past, I&S was created in 2009 to give both principals and teachers a larger role in determining internal assignments. There are a total of five stages to I&S (Diagram 1). At the beginning of each round of I&S, schools post vacancies on the I&S platform where every teacher in MPS can see and apply online if they would like to transfer. Each teacher applying for any transfer uploads a single resumé, with a suggested standardized format, that goes to all schools applied to by that teacher in that round. Once the posting round is closed, for each posting the system automatically sends interview invitations to the top five most senior applicants. If the posting garners more than five applicants, the principal, often supported by a school-based hiring committee, then has the option to send interview invitations to up to five other applicants, regardless of their seniority.[10] Seniority is measured by the first date of employment with MPS. Before submitting the application, applicants know how senior they are, but do not know how their seniority is *relative* to other applicants. After all the interviews are conducted, schools rank up to four applicants for whom they are willing to offer the position.[11] At an established date and time, email offers simultaneously go to the first-ranked applicant for each position. These applicants have 48 hours to accept one of

[10] Beginning in 2014, the automatic interview cutoff dropped to four most senior applicants and the maximum number of interviews declined to eight. In other words, schools can select up to four candidates (not five) regardless of their seniority if the posting garners more than four applicants.
[11] If a posting garners fewer than four applicants, the school has the option to not offer the position to anyone. From 2009 to 2013, if a posting garnered four or more applicants, schools were required to offer the position to an applicant even if they would prefer to make no offers and perhaps re-open the posting in later rounds. From 2014 onwards, schools are no longer required to make the offer even if they garner four or more applicants.

any offers made. For any positions where the first-ranked applicant declines or does not respond within 48 hours, a new set of email offers automatically goes to the second-ranked applicant. If offers are still not accepted, this process repeats a third and fourth time within round. An applicant who accepts an offer in an early set can forfeit it to accept an offer that arrives in a later set. If no ranked applicant has accepted an offer at the end of the fourth set of offers, the posting remains open and the school can repost the position again in a second round. Two rounds of the I&S process are conducted each year, each with applications, interviews and up to four offers.[12] Only after the I&S process can schools look for external applicants.

Given these transfer rules, the internal teacher labor market in this context is neither entirely deterministic nor completely decentralized. On the one hand, I&S still exhibits characteristics of a job-queuing process commonly observed in most American school districts. The more senior the applicant is relative to her competitors, the more probable she will successfully transfer just because she is more likely to automatically qualify for an interview. If the cost of search for teachers is not negligible, or if the cost of receiving a rejection is high, teachers may choose to wait and queue up for a more desirable job. On the other hand, the process gives MPS principals a high level of hiring autonomy and offers less senior teachers a chance to transfer. Starting in 2009, principals and their hiring committees can now interview and hire any applicant they desire. As such, both teacher

[12] For a more detailed description on the transfer rules of I&S over time, see Table A.1. Table A.2 tests the integrity of the rules.

and school preferences likely will both play a role in shaping the distribution of teachers in the MPS context.

III. Literature Review

Research on teacher mobility and the teacher labor market in the US has found that urban schools typically serve high concentrations of low-income, low-achieving, and students of color, and these schools tend to have lower teacher quality and a higher teacher turnover rate. Studies have found that teachers in urban schools tend to be less experienced (Clotfelter et al. 2005; Clotfelter et al. 2006; Rockoff 2004), and assigned to teach subjects that do not match their qualifications (Ingersoll 2003). In one of the largest studies of the distribution of teachers, Lankford et al. (2002) examines the variation in the average attributes of all public school teachers in the New York State from 1985-2000 and finds that urban schools not only have teachers with lower pre-service qualifications as measured by standardized test scores but also have a higher teacher turnover rate compared to suburban schools.

Teacher mobility may be one reason for this sorting pattern. Compared to teachers who remain, teachers who move often have better pre-service qualifications and tend to move away from hard-to-staff schools (Boyd et al. 2005; Goldhaber, Gross, and Player 2007). More importantly, teachers who leave also generally move to schools with lower concentrations of minority and disadvantaged students, and a higher level of student achievement (Boyd et al. 2005; Hanushek, Kain, Rivkin 2004; Hanushek, Rivkin 2007; Scafidi, Sjoquist, and Stinebrickner 2007). The distribution of principal quality and principal sorting behavior follow a similar pattern. A large number of studies have documented that schools serving disadvantaged students are more likely to have high rate of principal turnover and smaller principal applicant pools (Branch, Hanushek, & Rivkin, 2009; Gates, Ringel, Santibanez, Ross, and Chung, 2003). Loeb, Kalagrides, and Horng (2010) find that in Miami-Dade county schools serving disadvantaged students tend to have principals who have less work experience, less education, and graduated from less selective colleges and universities. Additionally, a series of follow-up surveys confirm that principals and assistant principals prefer to work in easier-to-serve schools with favorable working conditions.

In contexts where school leaders have some level of hiring autonomy, school preferences for teachers are equally likely to be influential in creating these transfer patterns. Qualitative studies suggest that the most sought after teacher attributes are communication skills, strong academic background, enthusiasm, interpersonal skills, flexibility, and ability to work in a team (Johnson and Roellke 1999). Béteille, Kalogrides, and Loeb (2012) examine schools in Miami-Dade County and find that high-achievement schools are able to attract and hire more effective teachers, assign new teachers to students in a more equitable fashion, and retain higher quality teachers more successfully than other schools. DeArmond, Gross, and Goldhaber (2010), however, argue that much of the difference in hiring outcomes among schools may come from schools' "local contexts," not from its approach to staffing. Recruiters in hard-to-staff schools are no more aggressive in their hiring style than recruiters in easier-to-serve schools. It was the local contexts such as school attractiveness that drove the different hiring outcomes.

Addressing the challenges concerning the distribution of teacher quality is difficult if one cannot separately identify how much a job transfer is due to teacher preferences or to school preferences. Due to data limitations, only a handful of studies can empirically disentangle teacher and school preferences. Boyd et al. (2011) uses applications-to-transfer data and work history data in New York City to disentangle the role of teacher preferences from the role of school preferences in explaining observed differences in transfer patterns. They find that teachers with stronger pre-service qualifications are more likely to request transfers. On the other hand, teachers with higher post-hire measures of quality and more experience are less likely to request transfers. For school preferences, they find that all schools exhibit preference for all types of quality measures. Boyd et al. (2013) combines the use of a game theoretic, two-sided matching model with method of simulated moments (MSM) estimation to study factors affecting the match of elementary teachers to their first jobs. They find that schools in some New York districts prefer teachers who have strong academic achievement and live close by and teachers prefer schools that are close geographically, suburban, and serving relatively advantaged students. Compared to non-white teachers, white teachers tend to prefer schools serving small proportions of minority students, a result we also document in our context.

Two recent works add both theoretical depth and new insights to the inner workings of the teacher labor market. Ahn (2015) analyzes North Carolina public school data and a survey of teachers' career plans and finds that teachers vary in their willingness to transfer and choose different search strategies at different stages of their careers. Mid-career teachers with high pre-service qualifications are the most eager to transfer while high-experience teachers are the least likely to transfer. Additionally, the author finds that new teachers typically start at a low-performing school before moving up to better schools as they gain experience. Teachers from low-performing schools search for high-performing schools about 90 percent of the time. Likewise, low-performing schools still search for high-experience teachers about 80 percent of the time even if they face the risk of being turned down. However, high-performing schools focus their search only on high-experience teachers. In the Florida setting, Feng and Sass (2016) find that where teachers are in the distribution of effectiveness can influence their transfer rate. Particularly, teachers at the top and the bottom of this distribution move at a higher rate than teachers with average effectiveness. In addition, there is evidence of assortative matching where more productive teachers are more likely to stay if their peers are productive.

Teacher preferences are not the only force driving these transfer patterns. Bonhomme et al. (2015) use administrative data containing every contract between teachers and a primary school in the Netherlands to show that job characteristics can affect teacher turnover not only through teacher preferences but also through their effect on access to other job opportunities. In other words, a teacher may remain in her job not only because she gets utility from doing so, but also because the current job attributes reduce her access to more attractive job opportunities. For example, teachers working in low-achieving schools may find it difficult to transfer to high-achieving schools or private schools. Any econometric model aimed at analyzing job transfers would need to account for the job characteristics of the origin-workplace as well. The authors find that teachers prefer schools with a smaller pupil-teacher ratio, higher average age of teachers, lower support-to-teaching staff ratio, lower proportion of disadvantaged and minority students, and more teaching hours (due to a salary schedule that varies with teaching hours). For the effect of the current job on other potential offers, they find that working in a school with a large proportion of disadvantaged and minority students increases access to an alternative school with similar student body attributes but lowers utility by decreasing access to schools with lower proportions of disadvantaged and minority students.

IV. Conceptual Framework

In this section, we present a conceptual framework of teacher and school preferences. A teacher would enter I&S only if she believes that a transfer to a new job will raise her utility. Likewise, a school would allocate its limited interview slots or make an offer only to a select group of applicants that will raise its utility. Since salaries are flat across schools, schools compete with nonmonetary characteristics such as school quality, location, and teacher and student demographics.[13] Similarly, teachers compete with their characteristics and potentially the characteristics of the schools they are currently teaching at. School and teacher attributes are not always completely revealed. For example, schools may be able to gauge the information on teacher ethnicity, effectiveness, and age at the resumé-screening stage but this information is likely to be more accurate when it is later revealed at the interview stage.

[13] There are, however, some cases where monetary incentives may be at play. From 2006 to 2010, some MPS schools adopted pay-for-performance regimes as part of Q Comp, a state law that was enacted in 2005. We exploit this school-level variation in the next section.

We start by modeling teacher preferences for schools. For simplicity, we omit time but we will add it back in the next section. In this internal labor market, there are *N* active teachers and *M* job postings. Within each posting, there can be up to *A* applicants, where $A \leq N$. The utility function of teacher *i* currently teaching at sending-school *j* is given by $u_{ij}(X_j, T_i, \epsilon_{ij})$, where X_j is a vector of the characteristics of school j, T_i is a vector of teacher i's characteristics, and ϵ_{ij} is an idiosyncratic difference among teachers' preferences. Assume that ϵ is i.i.d extreme value. Given this utility function and assuming that separability assumption holds, teacher *i* will apply to a potential receiving-school *k* from her sending-school *j* only if:

$$u_{ik}(X_k, T_i, \epsilon_{ik}) > u_{ij}(X_j, T_i, \epsilon_{ij})$$

$$f(X_k, T_i) + \epsilon_{ik} > f(X_j, T_i) + \epsilon_{ij}$$

$$\epsilon_{ik} > f\left(X_j, T_i\right) - f(X_k, T_i) + \epsilon_{ij}.$$

If ϵ_{ik} follows a logistic distribution, we can write the probability that teacher *i* applies to school *k* from school j as:

$$P\left(y_{ijk} = 1\right) = \frac{e^{f(X_j, T_i) - f(X_k, T_i) + \epsilon_{ij}}}{1 + e^{f(X_j T_i) - f(X_k, T_i) + \epsilon_{ij}}} \quad (1)$$

where y_{ijk} is a binary variable that takes value 1 if teacher *i* applies from school *j* to school *k* and 0 if otherwise.

This model incorporates not only the information about the attributes at the potential receiving-school (X_k) but also about teachers own characteristics (T_i) and the attributes of the sending-school (X_j). Note that the model allows the econometrician to include every active teacher in MPS at any given time the I&S was operating, not just those who participated in I&S. We include each teacher's own characteristics because there are reasons to believe that the interactions between teacher characteristics (that do not vary within teacher) and school characteristics (that vary by school) may produce differential effects on the utility received. For example, white teachers might receive more utility from working at a school serving high proportions of white students, a result documented in Boyd et al. (2013).

On the demand side, there are several ways to model school preferences due to the multi-stage nature of I&S. In the simplest case, we first consider the decision whether to move a candidate forward to the interview stage based on how her résumé compares to the minimum requirement to perform the job she applies to. Assume that the pool of candidates that receiving-school k can review for position s has A candidates. Let $v_{ik}(X_j, T_i, \varepsilon_{ik})$ be the utility that receiving-school k gains from employing teacher i from sending-school j, where X_j is a vector of school-level characteristics of school j, T_i is a vector of teacher i's characteristics, and ε_{ik} is an idiosyncratic difference among schools in their ability to assess the utility received from employing teacher i. Assuming that the separability assumption holds, school k will interview teacher i from school j only if:

$$v_{ik}(X_j, T_i, \varepsilon_{ik}) > R_{ks}$$

$$g(X_j, T_i) + \varepsilon_{ik} > R_{ks}$$

$$\varepsilon_{ik} > R_{ks} - g(X_j, T_i)$$

where R_{ks} is school k's reservation value of hiring a teacher for position s. In other words, school k will not offer an interview to teacher i if the utility it expects to gain from hiring her is less than R_{ks}, which will be estimated as school-position specific fixed effects. If ε_{ik} follows a logistic distribution, we can write the probability that school k selects teacher i as:

$$P(y_{ijk} = 1) = \frac{e^{R_{ks} - g(X_j, T_i)}}{1 + e^{R_{ks} - g(X_j, T_i)}} \qquad (2)$$

where y_{ijk} takes the value of 1 if school k decides to move teacher i from school j forward in the hiring process and takes value 0 if otherwise. We include sending-school characteristics because they may contain additional information that teacher characteristics do not. Essentially, we can test whether sending-school characteristics have any impact on the odds of being selected as done in Bonhomme et al. (2016). The sample includes every applicant for position *s*.[14]

As the hiring process passes through the interview stage, schools rank the candidates and decide who merit an offer. Model (2) will no longer be an accurate representation of this decision-making process. Instead, we can model how schools rank candidates by assuming that schools rank candidates by comparing them in a pairwise

[14] We are aware that some applicants are guaranteed to receive interview invitations based on their seniority. So in our estimation, we include an indicator of whether an applicant automatically qualifies by the seniority rule.

manner. When comparing teacher *i* from school *j* and teacher *z* from school *w*, receiving-school *k* will rank teacher *i* above teacher *z* only if:

$$v_{ik}(X_j, T_i, \varepsilon_{ik}) > v_{zk}(X_w, T_z, \varepsilon_{zk})$$

where X_w is a vector of the characteristics of school *w*, T_z is a vector of teacher *z* characteristics, and ε_{zk} is an idiosyncratic difference among schools in their ability to assess the utility received from employing teacher *z*. Let $y_{ak} = 1$ denote that school *k* most prefers candidate *a* out of all *A* potential candidates. In this case, it must be that:

$$v_{ak} \geq \max\{v_{1k}, \dots, v_{Ak}\}.$$

On the other hand, if $y_{ak} = 3$, this means that school k third-most prefers candidate a.

One way of using the information about which candidate schools rank first (prefer the most) out of *A* candidates in the interview pool is to model this decision as a multinomial logit model (McFadden 1973, 1974):

$$P(y_{ak} = 1) = P(v_{ak} \geq \max\{v_{1k}, \dots, v_{Ak}\})$$

$$= \frac{\exp(v_{ak})}{\sum_{q=1}^{A} \exp(v_{qk})} \qquad (3)$$

where $P(y_{ak} = 1)$ is the probability that teacher *a* ranks first by school k.

Alternatively, we can extract more information from school k's hiring decision by utilizing the post-interview ranking of candidates. For convenience, let r_{lk} be the candidate number that receives rank *l* by school k. The relation between r and y is given by:

$y_{ak} = l \leftrightarrow r_{lk} = a, \qquad\qquad \text{for all } a, l = 1, \dots, A.$

Therefore, $r_{3k} = 4$ means that school k ranks candidate number 4 third from the top. Likewise, $r_{Ak} = 1$ means that school *k* ranks candidate number 1 last. Using this notation, we obtain the rank-ordered logit model as specified in Beggs et al. (1981) and Chapman and Staelin (1982):

$$P(r_k) = P(v_{kr_{1k}} > v_{kr_{2k}} > \cdots > v_{kr_{Ak}})$$

$$= \prod_{a=1}^{A-1} \frac{\exp\left(v_{kr_{ak}}\right)}{\sum_{q=a}^{A} \exp\left(v_{kr_{qk}}\right)} \qquad (4)$$

where $P(r_k)$ is the probability of observing ranking r_k in a pool of *A* candidates at school k.

An important assumption underlying all of these models is that the choices that teachers and schools make must reveal their true preferences about one another. In some contexts, it can be beneficial not to rank the alternative that yields the highest utility first if the probability of actually being matched to that alternative is low. If strategic behaviors like this are in play, our estimates would be biased. On the teacher side, this violation is unlikely to be a concern due to three reasons. First, teachers do not rank schools. They simply need to apply. If they do not apply, they do not get to transfer. Thus, there is no incentive to be strategic if they want to transfer. Second, the cost of applying online is low. Even though some teachers may not think they have a chance of being hired at their most preferred school, the marginal cost of applying is close to zero since they only need to submit the standardized resumés that have already been created when they first created their I&S profiles. Additionally, there is no customization or cover letter needed. Third, teachers cannot completely gauge how likely they will be eligible for automatic interviews. They do not see how their seniority is ranked relative to their competitors before submitting the application. While there is a possibility that early-career teachers may be discouraged enough by this seniority privilege that it deters them from applying to schools that are commonly perceived as desirable, we address this issue by controlling for teacher experience when we estimate teacher preferences. On the school side, there is a possibility of strategic behavior, but it is likely to be limited. Some schools may choose not to send the offer to the most preferred candidate to avoid the possibility that this candidate may stall the process by taking the full 48-hour period and let the offer expire without accepting or declining. In this 48-hour period, it is possible that the second most preferred candidate may have already received and accepted offers from other schools. While we cannot rule out this concern, we argue that the fact that the district allows teachers to make up to two moves (accepting an offer) during each round of I&S can help alleviate this concern.

To empirically test whether there is evidence of strategic behavior on each side, we use the method in Hitsch et al. (2010) to split teachers and schools into ten quality bins and estimate the following regression:

$$1[apply]_{ik} = \sum_{b=1}^{10} \beta_b \cdot 1[quality_k = b] + \mu_i + \epsilon_{ik} \qquad (5)$$

where $1[apply]$ is a binary variable that takes value 1 if teacher i applies to school k; $1[quality_k = b]$ takes value 1 if the school that teacher *i* is considering whether to apply falls in to the b^{th} decile of school desirability as measured by the size of application pool and 0 if otherwise; μ is the teacher specific fixed effects; and ϵ is the random disturbance term. We estimate equation (5) for schools that fall into each quality bin separately using the information on whether teachers of different levels of teaching experience choose to apply there. We then plot the predicted probability of applying to a school against school quality separately by quintile of teacher experience in Panel A of Figure 2.2. Overall, we find evidence against strategic behavior since the predicted probability to apply mostly increases monotonically as school desirability rises regardless of applicant's quality.

On the school side, we estimate the following equation:

$$1[first]_{ik} = \sum_{b=1}^{10} \alpha_b \cdot 1[quality_i = b] + autoint_i + u_k + e_{ik} \qquad (6)$$

where 1[first] is a binary variable that takes value 1 if school *k* ranks teacher *i* first and 0 if otherwise*;* $[quality_i = b]$ takes value 1 if teacher i falls in to the b^{th} decile of effectiveness and 0 if otherwise; *autoint* is an indicator of whether teacher *i* automatically qualifies for an interview; *u* is the position specific fixed effects; and *e* is the random disturbance term. In Panel B of Figure 2, we plot the predicted probability of ranking applicant first against applicant effectiveness separately by quintile of school quality. The overall conclusion is similar to the test on the teacher side.

V. Data

The I&S data set contains information about every online action for every internal job posting in MPS from 2009 to 2015. Each record in this data set represents a potential match between applicant *i* from school *j* to position *s* at school *k* in year *y*. Each record contains the following information: job id, date of application, school id, round of application, applicant id, a ranking of seniority, whether applicant was selected for an interview, post-interview ranking, whether an offer was made, and whether the offer was accepted or rejected. Figures A.1 through A.4 present the basic descriptive statistics from the I&S data set. In 2014, I&S received roughly 3,000 applications from just under 500 unique applicants for 300 postings. Between 2009 and 2015, the number of postings as well as the number of applicants interested in transferring increased markedly, especially after 2013. On average, this translates into larger application pools and more options for the schools to choose from.

With the school and teacher identifiers from the I&S data set, we can establish a link with the administrative data sets to pull in additional information such as applicant attributes and school attributes, for both the sending and receiving schools, at multiple points in time. Teacher characteristics include, but are not limited to, gender, ethnicity, experience, and measures of effectiveness.[15] Table 2.1 provides the summary statistics across four types of MPS teachers between 2009 and 2015. The first column represents

[15] We measure teacher effectiveness with a year-specific composite score that pools all available effectiveness measures from math value-added, reading value-added, classroom-observations, and student surveys for the year of application. For more information on the measure of effectiveness at MPS, see Appendix B.

every MPS teachers who ever participated in I&S. The second column includes teachers who never participated in I&S. The third and the fourth columns split I&S participants in the second column into two groups: those whose applications were unsuccessful and successful.[16]

Between appliers and stayers (columns 1 and 2), there is little demographic difference. The most striking difference among them is that appliers tend to be significantly less effective. In terms of experience, these statistics confirm the results in Ahn (2015) that early to mid-career teachers are more likely to engage in search. One potential explanation for this pattern is the following. While seniority raises the probability of a successful transfer, the cumulative utility gain from a job transfer at the later stages of their careers may not be large enough to cover search costs. If the intuition from Ahn (2015) that teachers start from low-quality schools, move to high-quality schools as they gain experience, and improve their match quality every successive match is correct, then one would expect the utility gain for late-career teachers to be low, which is reflected in a smaller fraction of high-experience teachers among the movers than among the stayers. In terms of sending-school attributes, appliers tend to apply away from schools that have higher proportions of non-white teachers and non-white students, lower average teacher experience, more likely to be a priority school than the schools that stayers remain at. Appliers search for schools that serve lower proportions of non-white students, have greater average years of teacher experience, higher school quality ratings, and higher student

[16] We define successful application as receiving at least one offer in each year of I&S. Therefore, unsuccessful applicants are those that reveal an intent to transfer but cannot because they do not receive any offers.

achievement. These descriptive statistics confirm the patterns that are well documented in the teacher mobility literature. Among I&S applicants (columns 3 and 4), successful applicants tend to be younger, more likely to be white, significantly more effective, and more likely to be in their early to mid-career than unsuccessful applicants. They also tend to come from schools that have lower proportions of non-white teachers and non-white students, higher school quality ratings, and slightly higher student achievement in math.

Analysis of applications data at the school level reveals four facts about internal teacher transfer and student disadvantage at MPS.[17] First, schools serving the least-advantaged students get a third to a quarter as many applicants per opening as schools serving the most-advantaged students (Figure 2.3). Second, more-effective teachers are more likely to apply for transfer to schools serving more-advantaged students than schools serving less-advantaged students (Figure 2.4). The average applicant to the most-advantaged schools has been about 0.2 standard deviations more effective than the average applicant to the least-advantaged schools. Third, schools across the spectrum of student advantage hire applicants who are substantially more effective than their average applicant, despite the fact that they do not systematically know applicants' effectiveness ratings. This is evidence that all schools can recognize and do value effectiveness among applicants. The stable 0.2 standard deviation vertical distance between the blue and orange lines in Figure 2.4 expresses that all kinds of schools, regardless of student characteristics, hire transferring teachers who are substantially more effective than their average applicant.

[17] In figures 2.3 through 2.5, we classify schools using three measures of student disadvantage -- share students of color, share of students who are low-income as measured by receiving free or reduced price lunch, and share who are not proficient in reading.

Fourth, schools serving the most-advantaged students attract applicants with about 3 to 4 more years of MPS teaching experience than schools serving the least-advantaged. In the next section, we further explore these phenomena with regression models.

VI. Empirical Strategy

Estimating Teacher Preferences

To estimate teacher preferences for schools, we run the following logit model:

$$P_{ijky}(teacher\ i\ applies\ to\ school\ k \mid teacher\ i\ can\ apply\ to\ school\ k\ ) = \frac{e^{a}}{1+e^{a}}$$

where

$$a = a_0 + (X_{iky} - X_{ijy})a_1 + T_{iy}a_2 + \delta_y + round + \epsilon_{ijky} \qquad (T1)$$

In this model, teacher *i* applies to position *s* at school *k* out of all the options she can apply to that are available to her in year *y*. For every active MPS teacher, we construct her choice set based on the vacancies in each year and each teacher's licenses. For example, if a teacher is not licensed in elementary education in year *y*, postings at the elementary level in that year will not be included in her choice set. She cannot apply for those positions on I&S. As such, the choice set varies by teacher and by time. For teachers who did not apply, we assume they considered other alternatives available to them based on their licensing status and the licensing requirements for each posting but they chose to "apply" to their own position at their current school. On the contrary, we assume that teachers who participated in I&S considered a position at their current school but chose not to apply.

According to model T1, the probability that teacher *i* from school *j* in year *y* applies for a transfer to school k is a function of the differences between sending and receiving school characteristics $(X_{iky} - X_{ijy})$, teacher *i*'s characteristics T_{iy}, year indicator δ_y, round indicator, and a random error term ϵ_{ijky}. We include teacher characteristics such as experience, effectiveness ratings, and race to control for heterogeneous tastes that may vary among teachers. School characteristics include the share of students and teachers of color, the share of students proficient in reading, pupil-teacher ratio, pupil-teacher aide ratio, priority school status, pay-for-performance bonus amount, commute time from teacher *i*'s home, enrollment, and the percentage of English language learners, special education, and low income students.

To capture the effects of differential taste for each type of teacher, we introduce interactions between teacher characteristics and school characteristics:

$$a = a_0 + (X_{iky} - X_{ijy})a_1 + T_{iy}a_2 + T_{iy} * (X_{iky} - X_{ijy})a_3 + \delta_y + round + \epsilon_{ijky} \qquad (T2).$$

The interaction terms we explore are the interaction between teacher of color and percent of students of color, and the interactions between teacher effectiveness and pupil-teacher ratio, bonus amount, percent of student reading proficiency, and percent of students of color.

We also run specifications using conditional fixed effects logit models:

$$a = a_0 + (X_{iky} - X_{ijy})a_1 + \delta_y + \theta_i + round + \epsilon_{ijky} \qquad (T3)$$

$$a = a_0 + (X_{iky} - X_{ijy})a_1 + T_{iy}a_2 +$$

$$T_{iy} * (X_{iky} - X_{ijy})a_3 + \delta_y + \theta_i + round + \epsilon_{ijky} \qquad (T4)$$

where the teacher-specific fixed effects, θ_i, allows us to compare each teacher's choice to apply within herself over time, essentially differencing out all the time-invariant teacher-specific tastes that may vary among teachers. We can still recover the effects of differential taste from the interaction terms in model T4.

Estimating School Preferences

On the school side, we use a rank-ordered logistic model to analyze how school select candidates. Formally, our rank-ordered logit model is given by

$$P_{ijsy}\left(\begin{array}{c} ranking\ of\ teacher\ i \\ \mid teacher\ i\ was\ interviewed\ for\ posting\ s \end{array}\right) = \frac{e^b}{1+e^b}$$

where

$$b = b_o + T_{iy}b_1 + \eta_{sjy} + \epsilon_{ijsy} \qquad (S1).$$

Unlike in the case of teachers, the choice set for each school-posting pair is clearly defined. School j can only select candidates who apply to posting s at school j in year y in each round. This restriction is put in place with the posting fixed effects η_{sjy}. For each posting, school j ranks the candidates based on T_{iy}, teacher characteristics. To incorporate as much information as possible, our preferred specifications do not separately estimate school preferences at each stage of I&S.[18] The dependent variable is an ordered ranking from 1 to 4, where 1 is the highest ranking. Applicants who were screened out before the interview stage receive a censored ranking of zero. We omit automatic interview applicants who were not ranked in the top four because we cannot observe whether they were revealed preferred to other candidates who were not interviewed.[19] Since we expect some within-receiving-school serial correlations, we cluster the standard errors at the school level.

VII. Results

Teacher Preferences Results

According to Table 2.2, teachers prefer to apply to schools with lower proportions of non-white students and teachers, higher average teacher experience, higher enrollment, higher reading achievement, lower pupil-teacher ratio, higher expected pay-for-performance bonus amount, and schools that would reduce commute time from home. Overall, the magnitudes of the coefficients of school characteristics remain stable across specifications. On average, a one-minute increase in commute time and an additional student per teacher at receiving-school each reduces the probability of applying by one to two percent, confirming the earlier findings in the literature that these are among the most

---

[18] In the appendix, we provide estimation results from specifications that separately estimate school preferences by stage. For the resumé-screening stage, we run a logit regression with a dependent variable that is a binary choice whether to send an interview invitation to teacher *i*. For the post-interview stage, we rerun model S1 but only for the sample that progressed to the interview stage.

[19] An example of this problem is the following. If ten teachers apply, four automatically progress to the interview stage. Additional four will be chosen to be interviewed by the school. The remaining two are then screened out before the interview stage. We can say that the four applicants that the school chooses on its own are preferred to the two that are screened out. But we cannot say for certain that some of the four auto-interview candidates, if unranked or ranked below any of the four chosen by the school after the interviews are conducted, are preferred to those two.

important school characteristics that determine teacher mobility. In terms of student body, a 10 percent increase in the share of students of color reduces the probability of applying by about 2 percent while a 10 percent increase in the share of students proficient in reading increases the probability of applying by 1.5 percent. Controlling for observable school characteristics, teachers are still about 4 to 7 percent less likely to apply to a priority school.

In columns T2A, T2B, T4A, and T4B, we introduce interaction terms to investigate whether there is any heterogeneity in taste. We find that for every 10 percent increase in the share of students of color, teachers of color are 2.5 percent more likely to apply compared to white teachers. In terms of teacher effectiveness, there is no evidence that highly effective teachers prefer schools with high reading achievement more than low-effectiveness teachers do. There is, however, evidence that highly effective teachers prefer schools with higher share of white students. For a 10 percent increase in the share of students of color, a teacher whose composite effectiveness rating is 1 standard deviation greater than the district mean is about 1 to 2 percent less likely to apply than an average teacher. There is also evidence that more effective teachers prefer schools with greater expected pay-for-performance bonus amount.

The log likelihood estimates from Table 2.2 allow us to calculate the bundle of school characteristics or monetary incentive that would make any type of teachers indifferent about working at any pair of schools. To do this, we first calculate the marginal rates of substitution (MRS) among school characteristics.[20] Each cell in Table 2.3 is the

[20] Refer to Appendix C for the derivation of the marginal rates of substitution.

rate at which a teacher is ready to face a one unit increase in the column characteristic in exchange for the row characteristic while maintaining the same level of utility. Negative values indicate that the column characteristic must be forgone instead of received. For example, a priority school status requires a reduction of 47 percentage points in the share of students of color to keep utility constant. This reduction is not only impractically large but also unrealistic since student demographics are not controlled by school districts. Instead, we can fully or partially neutralize a priority school status by changing other school characteristics, even multiple at once. The district may decrease the number of students per teacher at that school by roughly 2.25 students and provide any applicant to that school a bonus of 116 dollars each year. Alternatively, the district can also just provide any applicant to that school a bonus of 332 dollars each year without altering the pupil-teacher ratio.

We propose two methods of calculating the dollar values of school characteristics. The first method is to derive the dollar values directly from the variation in the pay-for-performance bonus amount at schools that adopted Quality Compensation law (Q Comp) prior to 2011. The dollar value of each one unit of school characteristic is simply its coefficient divided by the coefficient of pay-for-performance amount. These dollar values are presented in the first column of Table 2.3. The second method is to use teacher preference for commute time that we estimate in Table 2.2 and the average value of travel time savings (VTTS) established in the transportation literature. Essentially, we multiply column 2 of Table 3 with the cost of time and the number of trips per year.[21]

To demonstrate how our calculations can be used in practice, we compute the changes in three policy levers required to create indifference about working at high and low achieving schools for different types of teachers. We begin by presenting the average school characteristics of two types of schools in Table 2.4. Schools in the bottom quintile of student reading proficiency have roughly half the total enrollments and serve three times as many students of color and low-income students as schools in the top quintile. Teachers at bottom-quintile schools have 6.5 fewer years of experience and about 9 fewer students than teachers at top-quintile schools. The pupil-teacher aide ratio at top quintile schools is roughly triple that of bottom quintile schools. Using these average school characteristics, we present the changes required at bottom-quintile schools to make them as desirable as top-quintile schools in Table 2.5. To attract an average MPS teacher working in top-quintile schools, the district must increase pay at bottom-quintile schools by $700 each year or reduce pupil-teacher ratio by 17. To attract a teacher whose composite effectiveness rating is 1 standard deviation greater than the mean, the pay must increase to $1,424 and the pupil-teacher ratio must decrease by 25. There is congruence between the two methods of calculating the dollar values of school attributes. The cost of change pupil-teacher ratio is orders of magnitude greater than the pay increase. This result does not imply that pupil-teacher ratio is not important for teachers. Our estimates in Table 2.2 indicate that it is one

[21] The nationwide average VTTS is estimated to be around $22.9 per hour between 2008 and 2010 (Belenky 2011).

of the most important school characteristics that drive teacher mobility. It is that bottom-quintile schools already have low pupil-teacher ratio. Further reducing becomes expensive and even impossible in some cases.

School Preferences Results

Throughout all stages of I&S, schools prefer candidates who are female, younger, white, holding an advanced degree, and effective (Table 2.6).[22] The probability that a teacher receives a higher ranking increases by 4 percent if she holds an advanced degree. This probability increases by about 6 to 7 percent for every 1 standard deviation increase in the composite effectiveness ratings. This result is evidence that schools can detect effective teachers even though they do not systematically know teachers' effectiveness ratings. Breaking down the effectiveness measures into its three components reveals that it is the SOEI ratings that primarily drive this result. Schools also prefer candidates who have no history of school-hopping. For two identical teachers, the one who stayed one additional year longer at her past schools would have a 2 percent increase in the probability of receiving a higher ranking. Experience is a positive trait, but the coefficients are not statistically significant at the 5 percent level. We do not find evidence that schools with a higher share of students of color prefer teachers of color more than white teachers.

Separately estimating school preferences by stage of I&S reveals that school preferences change with the information available to them. Controlling for who receives

---

[22] Our results remain stable with an alternative ranking where automatic interview applicants who were not ranked after the interviews are assigned the maximum ranking within the interview pool plus one (Table A.4)

automatic interview invitations, schools are more likely to move candidates who hold an advanced degree and have no history of school-hopping, and candidates who are younger, white, highly-effective, and in their late-career to the interview stage. Teachers whose composite effectiveness ratings are 1 standard deviation above the district mean are 3 percent more likely to receive an interview invitation.[23] Interestingly, it is the student survey component, rather than the SOEI or value-added scores, that drives this effect. Having an advanced degree increases the probability of receiving an interview invitation by about 6 percent. Having over 10 years of teaching experience increases the probability of receiving an interview invitation by over 15 percent. Controlling for years of experience, a one year longer tenure at past schools increases the probability of progressing to the interview stage by 3 percent, confirming our records from interviewing with MPS principals. However, it is rather surprising that the coefficients of some of the demographics are statistically significant. Information about them is either limited or missing at the resumé-screening stage. For age and race, the only way this information is revealed on the resumés is through names and date of birth.

After the interviews are conducted, schools start to pay less attention to the "signal" characteristics. The effects of experience and school-hopping history are indistinguishable from zero. The magnitude of the effects of holding an advanced degree decreases by more than half. On the other hand, the effects of age, being female, and effectiveness ratings all

[23] While the effectiveness ratings are never shared on resumés (or at any point in the I&S process). Their statistically significant positive effects indicate that school leaders prefer certain characteristics—unobserved by the econometrician—that are positively correlated with effectiveness ratings. Examples of these characteristics that may appear on the resumés are willingness to participate in additional professional development and the quality of their references.

increase in magnitude after the interview, possibly because more information about these characteristics is revealed in an in-person setting. An additional year of age reduces the probability obtaining a higher ranking by just under 1 percent. Being female now increases the probability of ranking higher by 4 percent. Having a composite effectiveness rating that is 1 standard deviation greater than the district mean increases the probability of ranking higher by 7 percent, almost double of its effect at the resumé-screening stage. There is also evidence that schools only exhibit preference for early offers candidates who tend to be high-performing and were pre-screened by the district only after the interview stage, not before.

We also find that being eligible for an automatic interview hurts the chance of being hired by schools *once a candidate finds herself in the interview stage*. All else equal, automatically qualifying for an interview decreases the probability of obtaining a higher ranking by almost 5 percent. We interpret this negative effect of the automatic interview eligibility as a potential loss of efficiency. On the school side, if the pool of application is sufficiently large, there may be some other high-quality candidates that the school may want to interview but cannot because the automatic interview candidates take up half of the available interview slots already. Theoretically, if a school has perfect information at the resumé-screening stage, it only needs to choose one applicant to interview. But, in this case, the interview stage would be superfluous. In the I&S context, schools do use information gathered through interviews, for better or worse, and the option to freely choose four or five other candidates may not be enough to guarantee that they can include the most desirable candidate in the interview pool. It is possible that more desirable candidates are not included in the interview pool because of incomplete information. On the teacher side, teachers who are highly desirable but are not eligible for an automatic interview would not have a chance to transfer to a school that would give them higher utility.

VIII. Conclusion

Estimating teacher and school preferences reveals that a large part of the sorting patterns we observe first stems from teacher preferences in the application stage. These patterns are then reinforced further by school preferences in the later stages of hiring. At the beginning of the hiring process, we find that teachers prefer schools with characteristics previously found to be desirable in many other contexts. As the hiring process progresses to where schools make decisions, we find that, when given the opportunity to make their own hiring decisions, schools general prefer candidates who are effective, holding an advanced degree, and not in their early-career. With these results in mind, it is not surprising why we observe the sorting patterns documented in many US school districts.

While these results are troubling since many of the school characteristics that teachers prefer are difficult to change in the short term, our study suggests that there is room for policy interventions that may be able to alter final hiring outcomes in a more immediate fashion. First, introducing an appropriate amount of monetary incentives at hard-to-staff schools may help attract more candidates and expand the choice set of these schools. We propose two methods to compute the appropriate amount of bonus to attract any type of teacher to any type of school. The size of the estimated bonus amount suggests that simply restoring the original pay-for-performance regime that was previously defeated district-wide at the end of 2010 is not sufficient to attract even an average teacher. Moreover, the amount of expected bonus can scale up considerably if the hard-to-staff schools need to attract teachers who need to relocate from far away. Second, supplying additional information about the applicants right at the resumé-screening stage can help reduce the likelihood that hard-to-staff schools miss out on quality candidates. Supplying more information can improve market efficiency, but whether it will improve or worsen the problem of inequitable distribution of teacher quality may depend on which schools receive the treatment. We leave this research question to future field experiments. Lastly, another potential policy lever is the transfer rule that currently favors seniority. We find evidence that schools exhibit strong distaste for candidates who automatically qualify for an interview. This confirms our discussions with school leaders that they would rather select every interviewee instead of letting the algorithm select half of them. With a large pool of applicants, the likelihood that schools can miss out on candidates that they would have preferred in the interview stage increases. If completely removing such rule is politically impossible, a more efficient automatic interview rule likely will have to take the size of the application pool into account.

We believe a number of next steps would improve this line of research. First, additional simulation exercises that take the estimated preferences as inputs may help shed light on how certain features of the teacher labor market affect teacher welfare, the efficiency of the teacher labor market, and the distribution of quality teachers. Estimating the effects of changing the seniority policy or the maximum number of candidates allowed in the interview pool may yield helpful insights for school districts and additional policy options that may prove more politically feasible than drastically changing school attributes or introducing pay-for-performance regimes. Second, given our preliminary findings that sometimes school leaders “miss out” on applicants who would have been highly preferred had they not been rejected at the resumé-screening stage, it is worth investigating further whether combining this kind of multi-stage hiring dataset with a value-of-a-hire prediction algorithm under a careful consideration of the appropriate counterfactuals can help improve hiring decisions. Similar to the way Kleinberg et al. (2017) shows how machine learning algorithms and econometrics can help improve bail decisions, it would be of great practical value to see how much prediction algorithms can help school leaders reduce the chance that they reject the wrong candidates before they have the chance to interview them.

Chapter 3:

Air Pollution and Academic Performance: Using a Natural Experiment in the Dallas-Fort Worth Metroplex

I. Introduction

Does air pollution affect academic performance? While air pollution has long been associated with a vast number of negative health outcomes, killing nearly 4,000 people daily in China and responsible for at least 200,000 premature mortalities in the United States each year, little is known about its impact on learning and skill development (Caiazzo et al., 2013, Rohde and Muller, 2015). From an economic point of view, school-aged children are at the heart of the process of human capital formation for any modern economy. Although most efforts are focused on directly improving the quality of education, the effectiveness of these interventions may be limited if children's capacity to learn is adversely affected by air pollution. Moreover, since air quality regulations have historically been tied almost exclusively to the impact of air pollution on premature mortalities, the gains from improving air quality may currently be underestimated if there is indeed a causal relationship between air pollution and skill development.

Only recently have researchers begun to discover the adverse effects of air pollution beyond health outcomes. Each pollutant may affect learning via different channels. For example, particulate matter and ground-level ozone (O3) aggravate asthmatic and respiratory symptoms in at-risk children, while nitrogen dioxide (NO_2) and carbon monoxide (CO) impair brain development, working memory, and neurological and cognitive functioning (Sunyer et al. 2015). If air pollution leads to reduced academic performance, adult outcomes of affected children may be at risk since they would be ranked below their peers, not because of their true capabilities, but because of an external factor that can be mitigated. To date, researchers have been able to convincingly link early-age learning and health outcomes to adult consequences (Chetty et al., 2011; Currie and Thomas, 2001; Currie et al., 2008). The potentially large negative externalities that have never been accounted for and the small but growing evidence in this emerging theme of research linking air pollution to important non-health outcomes warrant the need to empirically assess the relationship between air pollution and academic performance.

Estimating the impact of air pollution on any outcome remains a challenging task because the variation in the exposure to air pollution is rarely exogenous. Much of the variation in the levels of air pollution comes from human activities. Therefore, it is possible that using observational pollution data without any identification strategy can lead to finding a positive impact of air pollution on health and educational outcomes since much of the variation in air quality comes from economic activity, which can influence these other outcomes by other means. In addition to the nonrandom *causes* of the variation in air quality, some individuals may also determine their own *exposure* to air pollution. Chay and Greenstone (2005) demonstrate that individuals do in fact sort themselves based on their preference for clean air. Even in the same geographical area, optimizing individuals may still behave in a certain way to reduce their exposure to harmful pollutants by using air pollution index forecasts to plan on staying indoors on days with high levels of air pollution

(Aldy and Bind 2013; Neidell 2004; Neidell 2009). If left unaddressed, this selection effect can greatly understate the full welfare cost of air pollution.

To address these concerns, this paper exploits an exogenous change in the concentrations of air pollutants after the closure of Terminal E at Dallas/Fort Worth International Airport (DFW) after Delta Airlines (Delta) filed for bankruptcy in 2005 (DFW 2004). The basic idea is that the terminal closure would lead to a decline in enplanement, which would then lead to a decline in the concentrations of air pollutants in the areas around the airport. The identification assumption is that the closure affected student achievement only through air quality and that the closure was unrelated to other factors that could determine student achievements in the schools around DFW.[24] Using monitor-level air pollution data, I find that the terminal closure led to a significant decline in the levels of NO_2 and particulate matters up to 2.5 micrometer in size (PM-2.5) around DFW relative to the rest of Texas at the time of closure, but not for other pollutants (Figures 3.1 and 3.2).

The contribution of this paper is two-fold. First, it adds robust evidence to a small but growing literature on the effects of air pollution beyond health outcomes by exploiting a natural experiment. Using multiple comparison groups and multiple time periods, I employ difference-in-differences (DID) estimation to show that the test scores of test takers

[24] In theory, aircraft noise can also affect learning. A review by Klatte et al. (2013) indicates that mixed results are reported with respect to chronic effects on children's attention. While some studies have found a negative association between aircraft noise and reading performance, the authors note that insufficient control of student demographics and the existence of performance-enhancing effects of noise suggest that the harmful effects are likely small in magnitude. In this study, the closure of Terminal E affected both math and reading scores, separately. This finding suggests that aircraft noise is unlikely to be the main cause since, if it has any effect on test scores, it tends to affect only reading comprehension.

near the DFW area improved from 1.6 to 2.8 percent of a standard deviation relative to the rest of Texas and that this improvement is unique to the impact of the closure of Terminal E. Second, this paper provides new evidence that air pollution affects different types of test takers differently. Particularly, the effects are more pronounced among high-performing high school students and low-performing elementary school students.

A brief outline of this paper is as follows. In section II, I provide the background and literature review on air pollution research, and the context of the closure of Terminal E. In section III, I describe the data, the construction of key variables, and present the summary statistics of test takers and the levels of air pollution from the pre-closure and post-closure periods in the DFW area as well as in other parts of Texas. In section IV, I discuss the empirical challenges and describe my strategies to overcome them. In section V, I discuss my empirical results. Section VI concludes with a discussion of the implications of these results for education and environmental policy.

II. Background

How Air Pollution Affects Learning

The evidence from the fields of epidemiology and health economics points to the fact that air pollution is a threat to learning. Air pollution affects children's academic performance through three main mechanisms: student absenteeism due to respiratory illnesses, reduced cognitive functioning, and behavioral issues. Every major air pollutant commonly found in the ambient atmosphere has been documented to be positively associated with student absenteeism. Currie et al. (2009) finds that higher levels of CO significantly increase student absenteeism in Texas even in areas that passed the national air quality standards. In a study in Helsingki, Ponka (1990) finds that ambient levels of SO2 and NO_2 are positively associated with the student absenteeism and the number of upper respiratory infections observed in school-aged children. Particulate matters have been associated with absence rates in grade 1 to 6 in a study in Utah Valley (Ransom and Pope 1992). Gilliland, et al. (2001) find that high concentrations of ozone are correlated with student absenteeism. There has also been mounting evidence that asthmatic children miss school more than non-asthmatic children, with differences ranging from one day to one full week of school (Fowler, Davenport, Garg 1992). Some of the impact may also be irreversible. In the longest study of air pollution and children health in Southern California, Gauderman et al. (2004) find that teenagers in smoggy communities were nearly five times as likely to have clinically low lung function compared to those living in less polluted communities. Lung function deficits would lower the child's ability to recover from a cold and increase the risk of respiratory diseases.

Air pollution can also affect learning through mechanisms that do not involve respiratory diseases. Air pollutants such as NO_2 can affect brain development, memory, and cognitive and neurological functioning. Sunyer et al. (2015) find that children who attend schools in areas with lower levels of NO_2 experienced a 11.5 percent 12-month increase in working memory but those who attended schools in areas with higher levels of NO_2 experienced only a 7.4 percent 12-month increase.

Air pollution can also trigger symptoms or behaviors that inhibit learning at school. Bussing et al. (1995) found that children with severe asthma alone were nearly three times more likely to have severe behavioral problems than children without any chronic condition. In a cohort of 1,619 inner-city students in Rochester, New York, the parents reported that their children with persistent asthma scored worse on peer interactions and task orientation, and were more likely to exhibit shy and anxious behaviors compared to non-asthmatic children (Halterman et al. 2006). Given the evidence that air pollution can aggravate asthmatic symptoms, there are reasons to believe that it may lead to more behavioral problems at school for asthmatic students as well.

Overview of the Empirical Literature

The first strand of research on air pollution exploits quasi-experiments such as abrupt policy changes, random variations in traffic congestion, and closures of a major point-source polluter to estimate the impact of air pollution. In one of the most influential studies on air pollution and infant health, Chay and Greenstone (2003) uses non-attainment status as an instrumental variable to estimate the value of clean air and show that the instrumental variables estimates are much larger in magnitude. Moretti and Neidell (2009) uses boat traffic as an instrumental variable for air pollution. They argue that boat traffic near the ports of Los Angeles is highly correlated with air pollution but is largely determined by external factors beyond the control of the local residents. Schlenker and Walker (2016) exploits exogenous change in airport traffic congestion at eight largest US airports and finds that a one standard deviation increase in airport congestion increases the level of CO by 19 percent. In addition, they connect this finding to measures of health and found that taxi time, asthma and respiratory admissions are strongly related.

While researchers have extensively documented the relationship between air pollution and health outcomes, the existing literature on the relationship between air pollution and student achievement is extremely limited. Most papers focus on student absenteeism as the main outcome (Chen et al., 2000; Currie et al., 2009; Ponka, 1990; Gilliland et al., 2001; Ransom and Pope, 1992). Only recently have researchers begun to study the impact of air pollution on student achievement. Using individual level data and community air pollution data from the longitudinal respiratory health study of Southern California school children, Ham, Zweig, and Avol (2014) find economically significant impact of particulate pollution and NO_2 on student performance on the California standardized test scores. Lavy, Ebenstein, and Roth (2012) finds that short-term exposure to PM-2.5 and CO is associated with a significant decline in the probability of receiving the *Bagrut* high school matriculation certificate in Israel.

There are, however, mixed findings even for this new strand of research. For example, using a difference-in-difference-in-differences method, Currie et al. (2009) find a strong impact of CO on student absenteeism but finds mixed effects for particulate pollution in Texas schools during between 1996 and 2001. In contrast, Chen et al. (2000) find no impact of air pollution in general but a beneficial impact on student absenteeism in some specifications. Using a sudden closure of a steel mill in Utah Valley and temperature inversion, Ransom (1995) finds that particulate pollution had a strong impact on school absenteeism while CO had an unexpected beneficial impact even in their instrumental variable estimation models. Finally, Chen, et al. (2000) and Gilliland, et al. (2001) find beneficial effects of particulate pollution on school attendance. Confounding from other pollutants and weak identification strategies may drive these confusing results.

Background on the Closure of Terminal E

In 2005, DFW was the third busiest airport in the world enplanement and deplanement, handling over 700,000 operations in a year (ACI, 2005). In February of 2005, Delta Airlines closed its entire DFW hub located at Terminal E. The decision to close the hub was not due to the economic outlook of the Northern Texas economy. In fact, the county-level unemployment rates for counties surrounding DFW actually decreased slightly after the closure (BLS, 2005). Rather, the decision came at a time when Delta had already been struggling to maintain its own financial adequacy. Initially, closing the DFW hub and reallocating resources to the Atlanta hub was part of Delta's plan to avoid bankruptcy, which inevitably still happened in September of 2005.

As a result, Delta decreased daily departures at Terminal E from 254 to 18 and the number of leased gates went down from 28 to 4 (DFW, 2006). The closure led to approximately 11 percent year-over-year decline in the number of flights (DFW, 2006). It is possible that the closure could have led to job losses among parents of students attending schools near DFW. One could imagine that a closure-induced unemployment spell or income losses to the parents could indirectly hurt their children's academic performance. Tthe impact is likely to be limited since DFW board publically made a commitment that the affected DFW 1,700 employees would not be terminated (DFW, 2004). Additionally, there was even a slight unexpected increase in the 2006 passenger levels of 0.2% from 2005 (DFW, 2006).

III. Data

Air Pollution Data

To demonstrate that air pollution around DFW did in fact decline relative to the levels observed in other areas, I use hourly pollution concentrations at the monitor level from the Environmental Protection Agency (EPA) for four types of air pollutants: PM-2.5, CO, O3, and NO_2. For each monitor that was operating on a particular day, I calculate the daily pollution level by averaging the readings from 06:00 to 17:00 to approximate children's exposure to air pollution at school. To map daily pollution levels to an area, I follow the method described in Currie et al. (2009) and calculate the weighted mean pollution level using an inverse distance weighted formula to assign a higher weight to the monitors that are located closer to the area. The formula used to calculate the weighted daily mean at any area is the following:

$$P_a = \frac{\sum_{i=1}^{j} \frac{1}{d_{ia}^{\propto}} P_i}{\sum_{k=1}^{j} \frac{1}{d_{ka}^{\propto}}} \qquad (1)$$

where P_a is the mean pollution level at area a for the past *t* days before the test, *d* is the distance from each monitor to area a, P_i is the mean pollution level at monitor *i* for the past *t* days before the test, and *j* are the total number of monitors within a 10-mile radius, and α is assumed to be 2. For accuracy, I only include the data from monitors within 10 miles from each area.

Air Pollution Concentrations around the Closure of Terminal E

Using (1), I can also calculate the pollution levels around DFW over time and show that the closure of Terminal E was a pollution shock specific to DFW. According to Figure 3.1, most pollutants exhibited an overall long term decline in the concentrations, which could be due to improvements in aircraft and cars engine technology. The decline of NO_2 and CO appeared to have coincided with the closure of Terminal E. However, Figure 3.2 demonstrates that, relative to the rest of Texas, only NO_2 and PM-2.5 experienced a sharp decline right around the closure of Terminal E.[25] Even if technological advancements slowly lowered the levels of NO_2 and PM-2.5 everywhere over time, Figure 3.2 shows that the levels of NO_2 and PM-2.5 clearly fell more in DFW than in other areas at the time when Terminal E was closed. These graphical results support my identification strategy that exploits the exogenous change in the levels air pollution stemming from the closure of Terminal E.

Test Scores Data

I use the school-grade level assessment results from TAKS in Math and English Language Arts (ELA) as measures of academic performance. TAKS is a standardized state examination administered between February and April from 2003-2010 in Texas. TAKS

[25] There appears to be a break in the upward trend in CO right at the time of the closure as well.

was used in grade 3 to 10 to assess student attainment of reading, writing, mathematics, science, and social studies skills required under Texas education standards. These high-stake standardized assessments were used widely by public universities in Texas as one of the admissions criteria. Some universities used minimum TAKS scores requirement to determine admission eligibility. Hence, performing well on the TAKS was critical for students who intended to further pursue their education at the college level.

TAKS assessments are appropriate for this study because they were strictly administered on the same day across all public schools in Texas.[26] Unlike in California where high-stake assessments are often administered within testing windows, the Texas testing calendars allow for a more accurate estimation of the impact of air pollution on academic performance. This unique characteristic ensures that my estimates reflect only the effect of pollution levels measured before the assessments were administered. Using this feature, I can also test for the differences in the magnitude of the effects over time.

To construct the test scores dataset, I obtain the Texas testing calendars for all public schools in Texas during 2004-2007 from Texas Education Agency. Then, I select only the assessments that were strictly administered on a specific date throughout Texas and collected the relevant testing information such as testing dates, grades, and subjects for each of the selected assessments. For comparability with other empirical studies, I include only the information from the assessments that were in English. This set of testing

[26] I have contacted Texas Education Agency to ensure that retests and make-up sessions that may have taken place on other unofficial testing dates were rare and excluded from the main dataset used in this paper.

information is then merged with the assessment results and socioeconomic characteristics of test takers obtained from the TAKS Aggregate Data System. This paper will use raw score, passing rate, and commended performance rate as the three main outcome variables.[27] The TAKS Aggregate Data System also contains information regarding the characteristics and demographics of the test takers at the school-grade level for each assessment that took place, allowing me to control for potential confounders such as socioeconomic level, race, and gender composition among the students that may have varied around the same time that Terminal E was closed.

Summary Statistics

Although there are a number of major airports in Texas, no single comparison group is perfect. Therefore, this paper will present the results from using multiple comparison groups: all schools outside the DFW area, and all schools located within a 5-mile radius from Austin-Bergstrom International Airport (Austin), George Bush Intercontinental Airport (IAH), William P. Hobby Airport (Hobby) and San Antonio International Airport (San Antonio).

Table 3.1 shows the summary statistics of the school-grade observations from schools located within a 5-mile radius from DFW as well as from schools in other areas of Texas before and after the closure of Terminal E. I label observations that took place from 2006 to 2008 as being after the closure of Terminal E since the terminal closed in February

[27] TAKS Aggregate Data System also provides data on absence rate on test day. Due to many zero values, my main models may not be appropriate to estimate the effects on absences. See appendix A for absence results.

and most TAKS assessments were administered in February through April. There was too little time for the effect to materialize in 2005. In the DFW setting, test takers were predominantly white and were less economically disadvantaged than other areas. Even though most of the statistically different means of the variables in Table 3.1 were small in magnitude, their variations across time may cause biases if not controlled for. Therefore, I will control for the changes in the composition of test takers in all of my regression models.

IV. Identification Strategy

To compute the difference-in-differences (DID) estimates, I compare the test scores from the Texas Assessment of Knowledge and Skills (TAKS) from schools that are located near DFW to that of schools that are located elsewhere immediately before and after the closure of Terminal E. The DID specification is given by the following equation:

$$Y_{sg,d} = \beta_1 P_{sg,d} + \beta_2 T_s + \beta_3 TxP_s + D_{s,g} + D_{y,g} + D_d + D_m + \theta X_{sg,d} + \varepsilon_{sg,d}$$

(1)

where $Y_{sg,d}$ is the standardized TAKS scores for school s grade g on day d; $P_{sg,d}$ is an indicator for test dates that were after Terminal E closure; T_s is an indicator for being near DFW; $T_s x P_{sg,d}$ is an indicator for being near DFW and being after the closure of Terminal E; D_{sg} is a vector of school-grade fixed effects; $D_{y,g}$ is a vector of year-grade fixed effects; D_d is a vector of day-of-week fixed effects; D_m is a vector of month fixed effects; and $X_{sg,d}$ is a vector of school-grade level socioeconomic controls.

The year-grade fixed effects help account for the variation in the difficulty of the assessments given to each grade in each year. These fixed effects also help account for unobserved economic activities at the aggregate level that could influence both academic performance and pollution levels. The school-grade fixed effects allow me to compare the effects of air pollution on academic performance within each grade at each school over time. Additionally, these school-grade fixed effects can account for the variations in test scores and health outcomes that could come from age and control for other school-grade specific unobserved characteristics, such as the quality of air filter system in school buildings, the quality of education provided, the fixed geographical characteristics of schools, and the time of day and the intensity of physical education classes.

Although the school-grade fixed effects should account for much of the variation coming from the qualities and abilities among the test takers within the same school-grade unit, the composition of test takers may change over time due to reasons such as absence on test date and sorting (Table 3.1). As such, I include demographic controls in my regression models to account for potential omitted variable bias stemming from variations in the composition of test takers. Month fixed effects are included to control for the variation in seasonal weather patterns and pollen counts that may affect student health. Lastly, day-of-week fixed effects would capture the effects of fatigue.

The causal effect of the closure of Terminal E at DFW on $Y_{sg,d}$ is the estimated by β_3 regardless of any other trends in student achievement that could have occurred at the time when Terminal E was closed. All of my regression models are weighted either by the number of test takers and/or the distance away from the airport. Observations with a larger number of test takers and observations located closer to a nearby airport receive more weight to focus on population with more intense exposure to air pollution from the airport. Standard errors are clustered at the school level as I expect that there may be serial correlations in unobservables over time within schools.

V. Results

After conducting robustness checks, I find that the closure led to a 1.0 to 2.8 percent of a standard deviation increase in the TAKS assessments scores and a 3.7 to 7.3 percent increase in the proportion of students with commended performance (Table 3.2). In other words, if I compare the TAKS scores of test takers attending schools located near DFW to those of test takers attending schools located elsewhere immediately before Terminal E was closed, the former rose between 1 and 2.8 percent of a standard deviation faster than the latter a year after the terminal was closed.

The effects are also heterogeneous across my samples (Table 3.3). I find that the closure led to a larger percentage increase in the scores among high-school-age test takers than in the elementary-school-age test takers. While the effect on the passing rate is not distinguishable from zero for an average test taker, I find that, among elementary-school-age test takers, the closure led to a statistically significant 0.8 to 5.2 percent increase in the passing rate.

To further test the strength of my identification strategy, I conduct two additional robustness checks. First, I add unit-specific time trend variables to (1). So, (1) becomes

$$Y_{sg,d} = \beta_1 P_{sg,d} + \beta_2 T_s + \beta_3 TxP_s + \beta_4 \tau_{\mathrm{sg,d}}$$

$$+ D_{s,g} + D_{y,g} + D_d + D_m + \theta X_{sg,d} + \varepsilon_{sg,d} \quad (2)$$

where τ is an interaction between treatment, which is being near DFW, and a time trend variable. This specification tests whether the parallel assumption is violated. That is, I allow each area to have its own linear time trend. I find that β_4 is not statistically different from zero and the estimates of β_3 on the TAKS scores remain mostly the same and the estimates on the TAKS passing rate continue to be statistically insignificant (Table 3.4).While the estimates on the commended performance rate declined considerably for all comparison groups, almost all are still statistically significant. The second test uses a placebo treatment variable. Instead of using 2006 to indicate post-closure status, I use 2004 to estimate (1) instead. Theoretically, I should not be able to detect any effect if the terminal did not close until 2005. As expected, all estimates except one from the Austin sample are statistically and economically insignificant as shown in Table 3.5.

VI. Conclusion

The evidence presented in this paper suggests that the closure of Terminal E, which improved the air quality around the schools located near DFW, led to a small but statistically significant increase in the achievement of students near DFW. Using terminal closure as an exogenous treatment of improved air quality, I employ DID estimation and estimate the effect of the closure to be between 1.0 to 2.8 percent of a standard deviation increase in the TAKS assessments scores and a 3.7 to 7.3 percent increase in the proportion of students with commended performance. While the effects are larger among students at the higher end of the score distribution, I find that effects on the lower end of the score distribution are also present among elementary school students. For these younger test takers, the closure led to between a 0.8 to 5.2 percent increase in the TAKS passing rate. These results suggest that air pollution affects students heterogeneously. Its effects are larger on the high performing older students and on the low performing younger students.

The policy implications are three-fold. First, policymakers may need to factor air pollution into the decision of whether to allow new schools, kindergartens, or daycare facilities to be built near areas such as busy intersections, airports, and highways where the level of NO_2 tends to be high due to burning of fuel. This paper shows that the closure of an airport terminal does affect student achievement in the areas close to the airport. This result is consistent with earlier findings on the impact of air pollution on hospital admissions in the areas near airports in California (Schlenker and Walker, 2016). For regulators of airports, reducing aircraft idling time and flight delays can be beneficial for the lives of students around the area.

Second, the findings in this paper stress the need to further promote EPA's Indoor Air Quality (IAQ) management programs for more schools to adopt EPA's recommended best practices and educate students and their parents about the negative effects of air pollution on students' health at school and on academic performance. The effort to promote IAQ may especially be needed in schools with high proportions of economically disadvantaged students as this paper shows that the impact is disproportionately larger for these groups of students. In addition, the EPA can also encourage more schools to subscribe to its air quality index forecast program. With the IAQ program and air pollution forecasts available to them, more schools will be able to take preventative measures based on the air pollution forecasts and make certain that they advise teachers to prevent asthmatic students or students with acute respiratory symptoms from doing strenuous activities outdoors on days with high levels of NO_2.

Third, EPA may need to revisit the national standard for NO_2. While the levels of NO_2 in the US have declined dramatically in the past three decades, there is not been scientific evidence to ensure that long-term exposure to low levels of NO_2 is safe or that the costs of lowering the standards further exceed the benefits. While the current levels may no longer be the primary cause of premature mortality, this paper adds to the small but growing literature which shows that long-term exposure to low levels of NO_2 can create a negative externality to the society through reduced student achievement.

Illustrations

Table 1.1: Summary statistics of dependent variables

Frequency of Monthly Service Requests in a Census Tract	2005-2009			2010-2014			
	Mean	SD	Max	Mean	SD	Max	t-statistic
Sewer blockage and street light outage	9.50	12.1	597	8.06	10.9	750	-30.3
Electricity outage and heating complaints	10.9	22.4	493	10.5	25.8	3638	-4.22
N	108,577			133,334			

Table 1.2: Summary statistics of proxy variables and instrumental variables

	2005-2009		2010-2014		
Heterogeneity Measures	Mean	SD	Mean	SD	t-statistic
Racial Fractionalization	0.443	0.197	0.467	0.195	29.6
Language Fractionalization	0.507	0.151	0.518	0.146	17. 9
Gini Coefficient	0.442	0.067	0.451	0.064	32.4
N	108,577		127,048		

	Decennial Census 2000		
Instrumental Variables	Mean	SD	Max
Fraction of households earning more than $200,000 in 2000	0.026	0.049	0.639
Fraction of workers employed in the public sector in 2000	0.0476	0.0326	1
Fraction of workers employed in the information industry in 2000	0.0450	0.0332	1

Table 1.3: Summary statistics of control variables at the census tract level

Socioeconomic Demographic Variables	2005-2009		2010-2014		
	Mean	SD	Mean	SD	t-statistic
Population	4,207	2,154	3,953	2,167	-28.8
Number of Households	1,560	1,005	1,466	969	-23.1
Mean Household Income	71,624	41,652	78,638	44,911	39.06
Median Household Income	53,830	26,342	58,540	28,626	41.30
% population under 20 years of age	25.3	8.7	23.8	8.0	-44.2
% population aged between 21 and 40	30.8	8.4	31.6	8.9	22.6
% population aged between 41 and 65	31.5	5.9	31.8	6.0	12.7
% population aged over 66	12.3	6.3	12.7	6.5	14.8
% White population	46.4	31.5	43.5	30.3	-22.9
% Black population	25.3	30.8	26.1	30.8	6.8
% American Indian population	0.4	0.9	0.4	0.9	6.4
% Asian population	11.8	14.9	13.3	16.3	23.6
% Hawaiian or Pacific Islander population	0.03	0.2	0.04	0.3	9.8
% some other race population[28]	14.1	16.0	13.6	15.5	-7.3
% population in civilian labor force	57.2	9.5	56.6	9.7	-16.3
% population unemployed	8.4	4.8	10.6	5.8	99.1
% married households with at least one child	18.2	11.5	17.7	9.5	-10.9
% owner occupied houses	36.9	25.14	36.5	25.18	-4.1
% population 25 years or older with Bachelor's Degree or higher	30.7	20.1	32.5	20.3	21.9
% of foreign born population	36.0	15.8	37.5	15.4	23.1
% of households where English is not spoken at	46.4	22.7	47.6	22.6	13.6
% of households in the same house last year	89.3	6.37	89.4	6.69	6.3
N	108,577		127,048		

[28] The ACS does not categorize Hispanic population. Hispanic population, along with residents who claim more than one races, are included in the some other race population variable.

Table 1.4: OLS and Fixed-Effects Results

	(1)	(2)	(3)	(4)
Gini Coefficient	-2.096***	-0.925***	-0.449**	-0.469**
	(0.148)	(0.161)	(0.191)	(0.186)
Racial Fractionalization	-0.311***	-0.071	-0.103	-0.115
	(0.061)	(0.074)	(0.089)	(0.104)
Linguistic Fractionalization	0.111	0.0154	-0.092	-0.150
	(0.074)	(0.094)	(0.096)	(0.114)
Log Population	-0.472***	-0.411***	-0.726***	-0.711***
	(0.028)	(0.051)	(0.056)	(0.061)
Log Median Household Income	0.181***	-0.054	-0.117***	-0.095***
	(0.025)	(0.044)	(0.04)	(0.042)
Observations	228,501	228,501	228,501	228,501
R-squared	0.173	0.300	0.123	0.125
F	141.5	11.41	194.0	59.89
Tract FE	No	No	Yes	Yes
Month-Year FE	No	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Note: Robust standard errors are clustered by census tract and year. The dependent variable is the inverse hyperbolic sine transformed per capita frequency of service requests about sewer blockage and street light outage. Racial and linguistic fractionalization measures are calculated using the reverse the reverse Herfindahl–Hirschman formula. Controls are the time-varying variables listed in Table 3 and the fractions of workers in employed in each industry. *p < .10. **p < 0.05. ***p < .01.

Table 1.5: 2SLS Results

	(1)	(2)	(3)	(4)	(5)
Gini Coefficient	-2.335***	-4.927**	-2.474***	-1.700**	-2.977***
	(0.721)	(2.480)	(0.731)	(0.793)	(0.782)
Racial Fractionalization	-0.002	0.025	-0.001	-0.078	0.065
	(0.084)	(0.091)	(0.084)	(0.101)	(0.095)
Linguistic Fractionalization	0.060	0.030	0.058	0.208	-0.025
	(0.104)	(0.113)	(0.104)	(0.124)	(0.108)
Log Population	-0.366***	-0.282***	-0.364***	-0.260***	-0.389***
	(0.053)	(0.072)	(0.053)	(0.051)	(0.057)
Log Median Household Income	-0.173**	-0.414**	-0.186**	-0.117	-0.247**
	(0.085)	(0.235)	(0.085)	(0.089)	(0.098)
Observations	228,226	228,226	228,226	104,519	123,707
R-squared	0.221	0.177	0.22	0.205	0.239
F-stat	35.99	30.49	35.74	26.67	31.72
First-stage F-stat	191.7	25.54	103.4	60.83	68.02
Month-Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
2000 Top Income IV	Yes	No	Yes	Yes	Yes
2000 Public Sector IV	No	Yes	Yes	Yes	Yes
Sargan-Hansen p-value	-	-	0.349	0.845	0.356
Sample	Full	Full	Full	ACS1	ACS2

Note: Robust standard errors are clustered by census tract and year. The dependent variable is the inverse hyperbolic sine transformed per capita frequency of service requests about sewer blockage and street light outage. Racial and linguistic fractionalization measures are calculated using the reverse the reverse Herfindahl–Hirschman formula. In column 1, the Gini coefficient is instrumented with the fraction of households with income greater than 200,000 dollars in 2000. In column 2, the Gini coefficient is instrumented with the fraction of working population employed in the public sector in 2000. Columns 3 through 5 include both instruments. Controls are the time-varying variables listed in Table 3 and the fractions of workers in employed in each industry. *p < .10. **p < 0.05. ***p < .01.

Table 1.6: 2SLS Results with Alternative Request Types

	(1)	(2)	(3)	(4)
Gini Coefficient	-0.052	-1.831***	-1.478*	-2.042***
	(0.164)	(0.611)	(0.799)	(0.610)
Racial Fractionalization	-0.057	0.008	-0.050	0.050
	(0.100)	(0.090)	(0.109)	(0.100)
Linguistic Fractionalization	-0.089	-0.002	0.050	-0.036
	(0.119)	(0.111)	(0.127)	(0.124)
Log Population	-0.575***	-0.087*	0.093*	0.090*
	(0.077)	(0.044)	(0.049)	(0.045)
Log Median Household Income	-0.108**	-0.372***	-0.362***	-0.371***
	(0.045)	(0.074)	(0.091)	(0.078)
Observations	195886	195850	87428	108422
R-squared	0.480	0.220	0.258	0.199
F-stat	132.6	30.64	32.93	21.88
First-stage F-stat		119.2	56.50	73.21
Tract FE	Yes	No	No	No
Month-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
2000 Top Income IV	No	Yes	Yes	Yes
2000 Information Sector IV	No	Yes	Yes	Yes
Sargan-Hansen p-value	-	0.230	0.484	0.271
Sample	Full	Full	ACS1	ACS2

Note: Robust standard errors are clustered by census tract and year. The dependent is the inverse hyperbolic sine transformed frequency of service requests about power outage and heating complains. Racial and linguistic fractionalization measures are calculated using the reverse the reverse Herfindahl–Hirschman formula. In columns 2 through 3, the Gini coefficient is instrumented with the fraction of households with income greater than 200,000 dollars in 2000 and the fraction of working population employed in the information industry in 2000. Controls are the time-varying variables listed in Table 3 and the fractions of workers in employed in each industry. *p < .10. **p < 0.05. ***p < 0.01.

Figure 1.1: Service Requests (height) and Income Inequality (shade) from 2010-2014

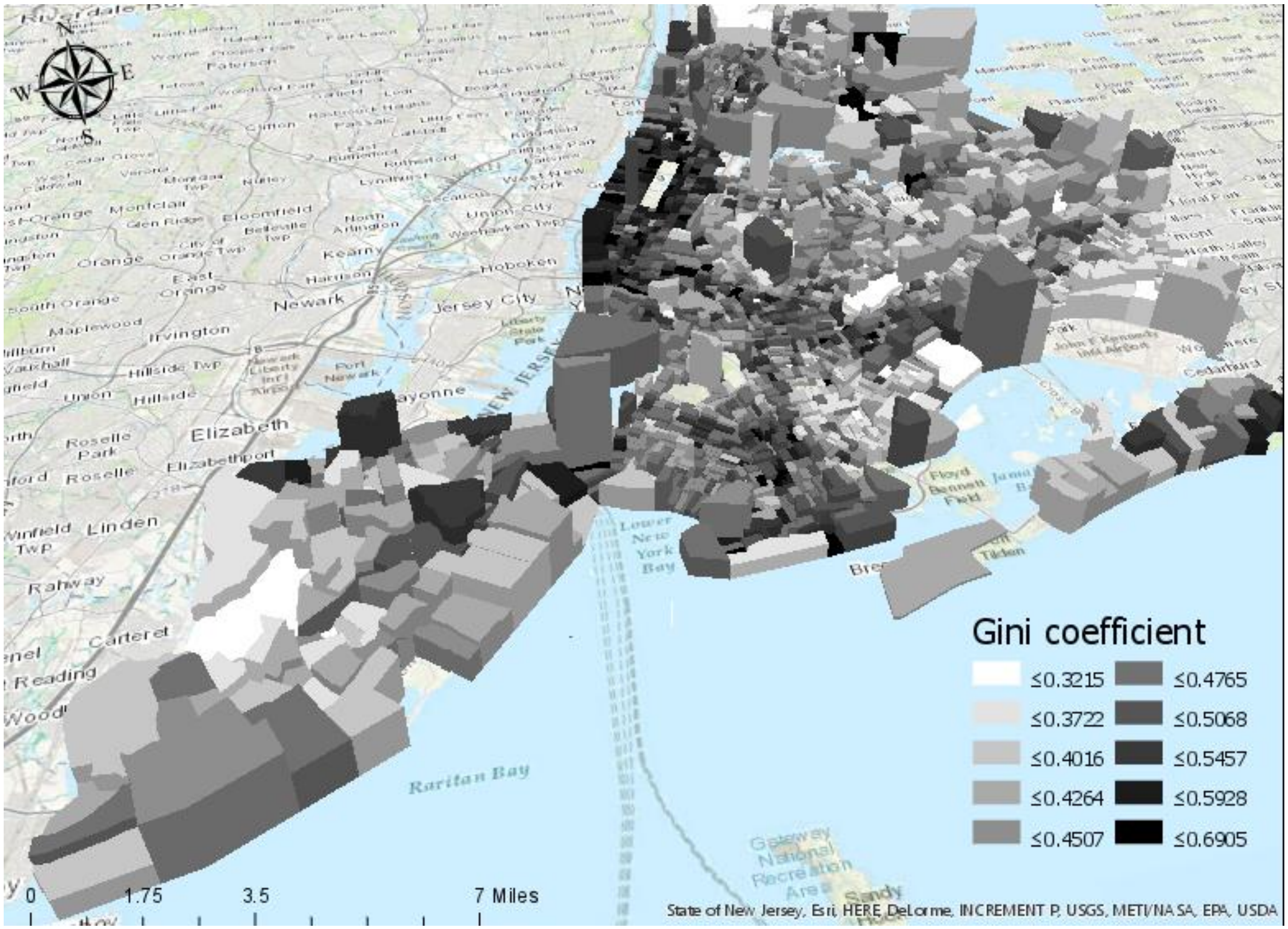


Figure 1.2: Service Requests (height) and Racial Fractionalization (shade) from 2010-2014

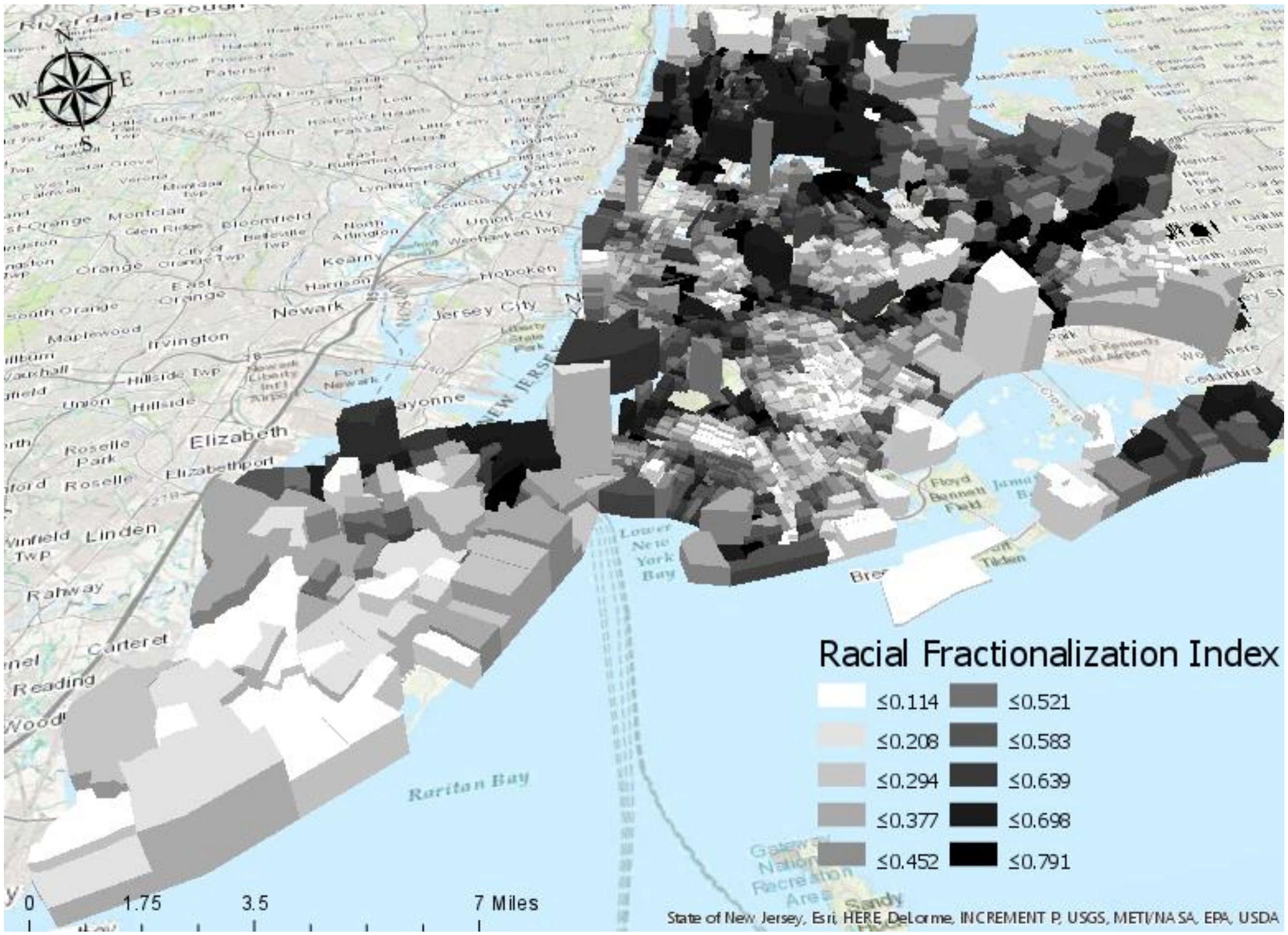


Figure 1.3: Service Requests (height) and Language Fractionalization (shade) from 2010-2014

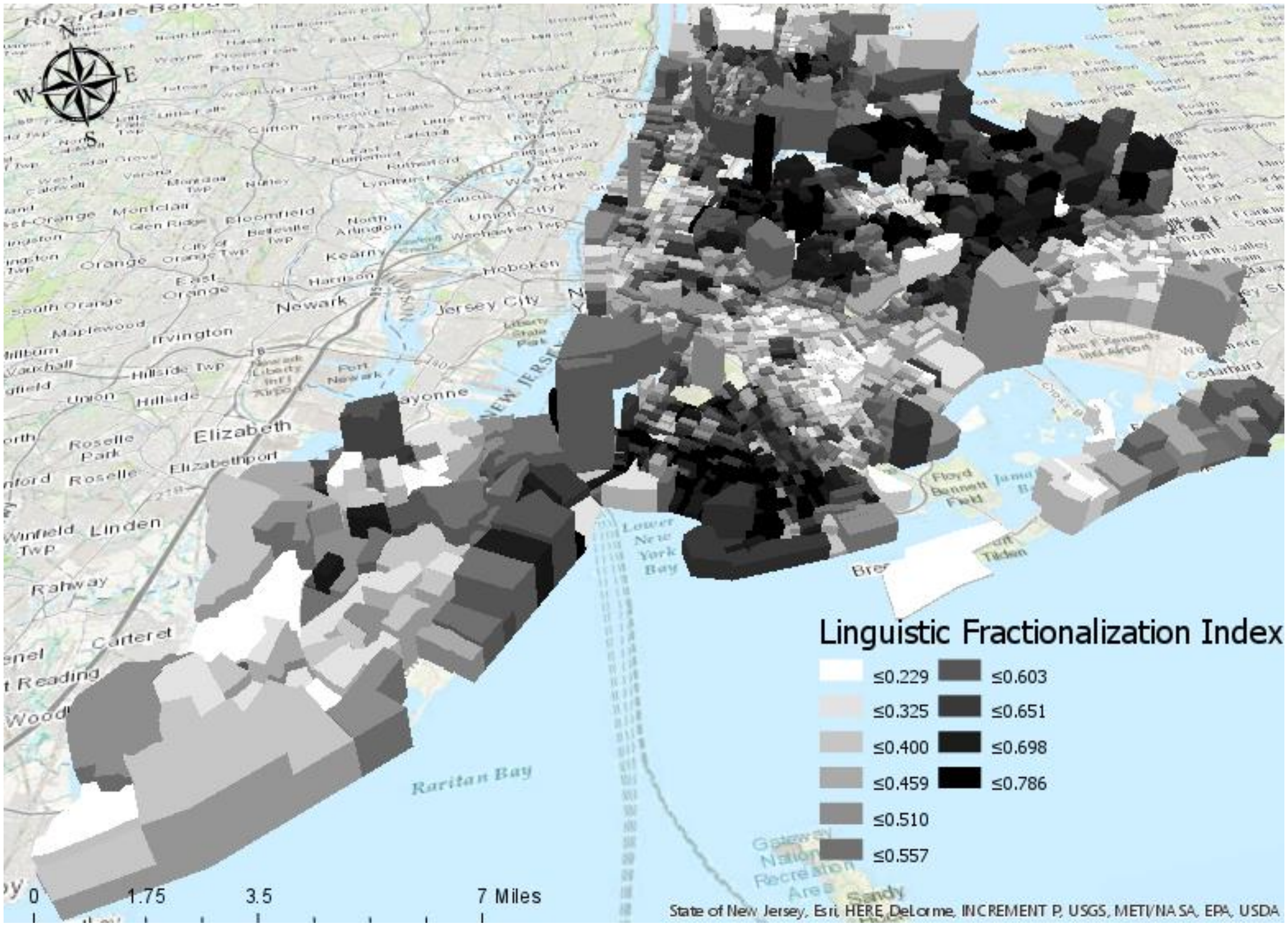


Figure 1.4: Correlation Heat Map of Average Service Request Completion Time by Census Tract and Demographics

Median Income										
-0.228	% of Black population									
0.451	-0.764	% of White population								
-0.161	-0.067	-0.276	% Foreign Born							
-0.058	-0.008	-0.020	-0.024	Fixing Time: Street Light Outage						
0.000	-0.008	0.010	-0.004	0.006	Fixing Time: Water Quality					
0.020	-0.018	0.026	0.047	-0.051	-0.004	Fixing Time: Electrical Outage				
0.056	-0.001	0.001	0.059	-0.082	0.001	0.097	Fixing Time: Heat/AC			
-0.049	-0.029	-0.012	0.045	-0.016	-0.002	0.023	0.034	Fixing Time: Sewer		
0.081	-0.086	0.125	-0.072	-0.043	-0.002	0.050	0.024	0.022	Fixing Time: Noise Complaint	
0.011	0.030	-0.002	0.025	0.000	0.000	-0.010	-0.021	0.017	0.017	Fixing Time: Potholes

Note: Font size indicates the strength of the correlation. Red indicates a negative correlation. Blue indicates a positive correlation.

Figure 1.5: Relationship between Service Requests and Instruments

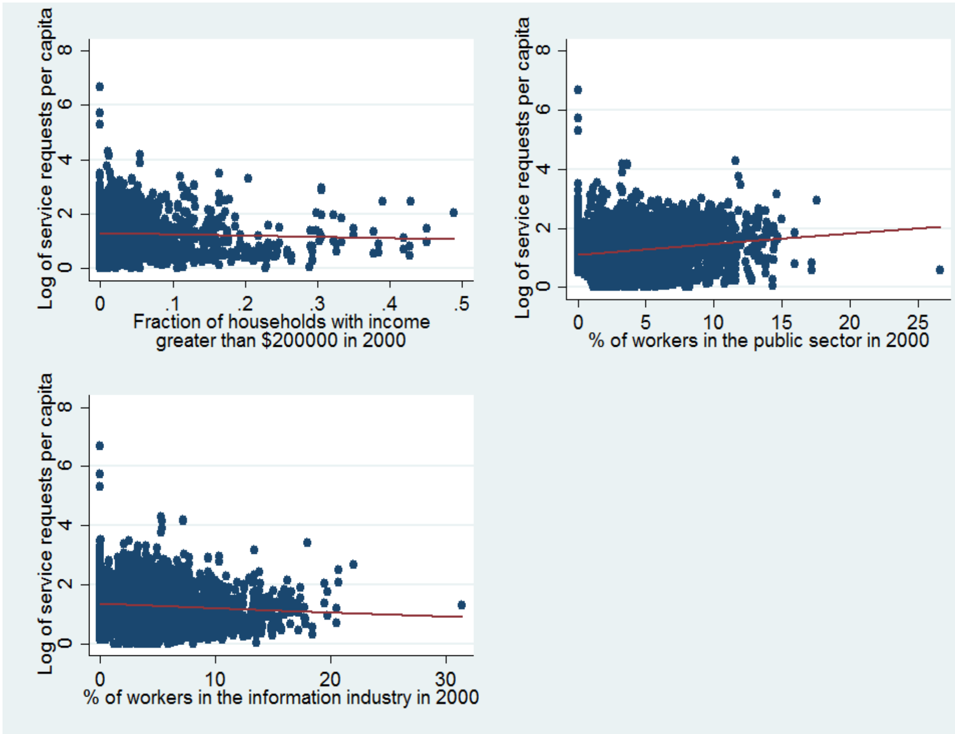


Figure 1.6: Relationship between Gini Coefficient and Instruments

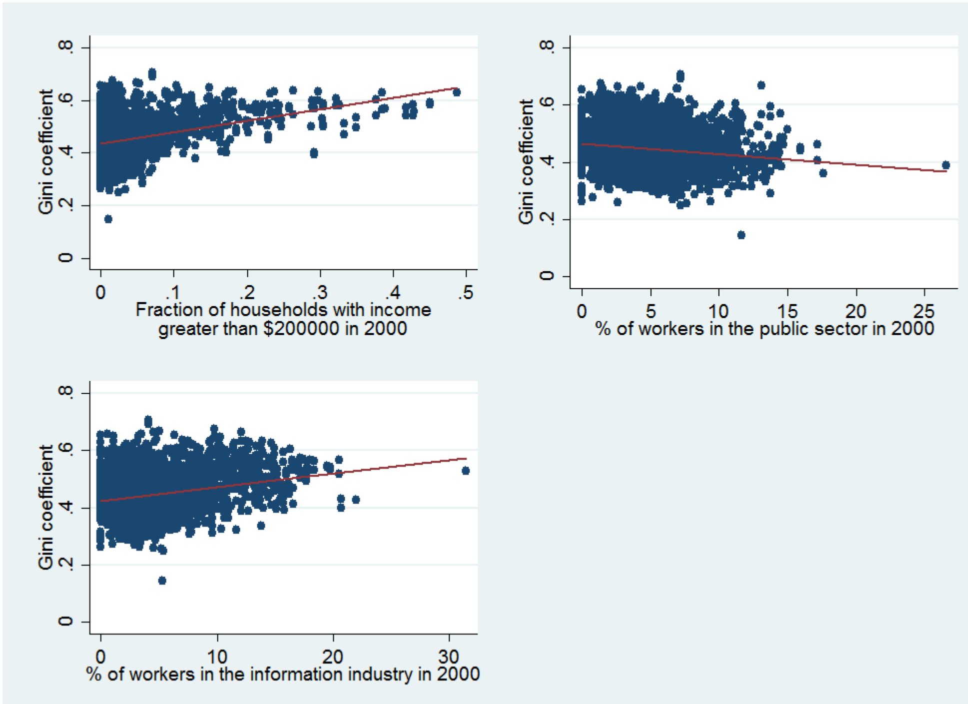


Diagram 1: Stages of Interview & Select

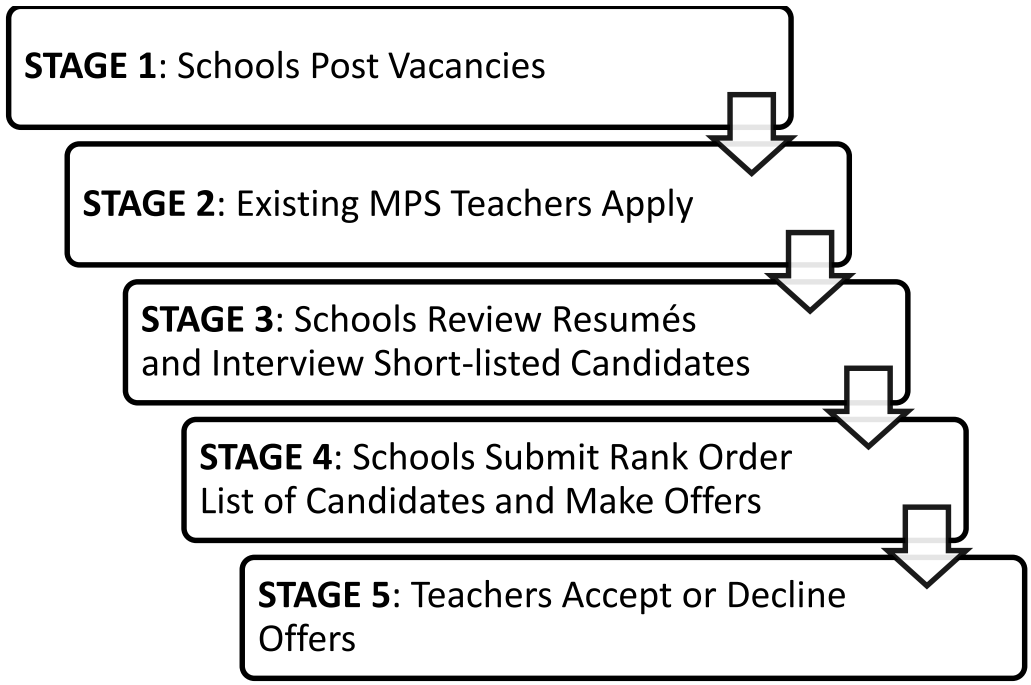


Table 2.1: I&S Applicants and Non-applicants from 2009-2015

	Movers	Stayers	uccessful Applicants	Unsuccessful Applicants
Female	0.740	0.747	0.766	0.710
Age	42.537	43.038	40.144	45.404
Teacher of color	0.194	0.181	0.175	0.217
Advanced degree	0.445	0.436	0.422	0.471
Effectiveness (z-score)				
Composite	-0.177	-0.005	-0.063	-0.328
Value-added	-0.061	0.002	-0.037	-0.094
SOEI	-0.199	0.009	-0.082	-0.353
Student Surveys	-0.070	-0.002	-0.0004	-0.171
Experience				
Early Career (1-3 years)	0.160	0.119	0.186	0.129
Mid Career (4-10 years)	0.180	0.140	0.201	0.154
Late Career (10+ years)	0.492	0.512	0.412	0.588
Sending-School Attributes				
Commute time (minutes)	18.4	16.9	18.2	18.6
% teachers of color	15.3	14.6	15.2	15.4
% students of color	78.5	71.2	77.7	79.6
Avg. years of teacher experience	12.8	13.8	12.7	12.8
% students proficient in reading	42.4	47.1	42.7	42.1
% FRPL	78.6	69.5	77.8	79.6
% ELL	27.2	25.9	27.0	27.4
% SPED	20.7	18.2	20.6	20.9
Total enrollment	615	706	619	611
Pupil-teacher ratio	14.6	15.7	14.8	14.4
Pupil-teacher aide ratio	42.6	51.4	43.6	41.4
Priority status	0.130	0.074	0.131	0.128
Receiving-School Attributes				
Commute time (minutes)	19.3		19.2	19.5
% teachers of color	15.8		15.6	15.9
% students of color	75.3		75.2	75.3
Avg. years of teacher experience	13.2		13.1	13.2
% students proficient in reading	46.5		46.6	46.4
% FRPL	74.2		74.0	74.3
% ELL	27.6		27.7	27.5
% SPED	17.3		17.2	17.3
Total enrollment	643		648	637
Pupil-teacher ratio	15.4		15.5	15.4
Pupil-teacher aide ratio	46.9		47.2	46.5
Priority status	0.126		0.124	0.128
Max Observations	2,934	23,006	1,592	1,341

Table 2.2: Teacher Preferences Results

		(T1A)	(T1B)	(T2A)	(T2B)	(T3)	(T4A)	(T4B)
Differences in school attributes (receiving– sending)	Avg. bonus amount	0.000712***	0.000722***	0.000703***	0.000708***	0.000861***	0.000852***	0.000847***
		(0.000210)	(0.000210)	(0.000210)	(0.000210)	(0.000060)	(0.000061)	(0.000061)
	Commute time	-0.041692***	-0.041258***	-0.040347***	-0.040020***	-0.073102***	-0.072061***	-0.072391***
		(0.003302)	(0.003309)	(0.003327)	(0.003370)	(0.003655)	(0.003733)	(0.003708)
	Priority status	-0.122747	-0.128381	-0.127542	-0.130664	-0.297106***	-0.315122***	-0.302929***
		(0.091040)	(0.092408)	(0.091755)	(0.092022)	(0.031961)	(0.032310)	(0.032085)
	Avg. teacher experience	0.009898	0.009772	0.008655	0.008899	0.027688***	0.026720***	0.027009***
		(0.012426)	(0.012468)	(0.012504)	(0.012486)	(0.004302)	(0.004349)	(0.004344)
	% students of color	-0.005787	-0.005788	-0.008638*	-0.008362*	-0.005741***	-0.008926***	-0.008563***
		(0.004549)	(0.004571)	(0.004686)	(0.004632)	(0.001950)	(0.002062)	(0.002057)
	% teachers of color	-0.000427	-0.000225	-0.000383	-0.000325	-0.002757*	-0.002537	-0.002677*
		(0.005335)	(0.005376)	(0.005339)	(0.005351)	(0.001589)	(0.001607)	(0.001600)
	% proficient in reading	0.006656**	0.006480**	0.006969**	0.006678**	0.006156***	0.006518***	0.006484***
		(0.003240)	(0.003251)	(0.003160)	(0.003248)	(0.000925)	(0.000919)	(0.000933)
	Enrollment	0.000299**	0.000291**	0.000313**	0.000301**	0.000502***	0.000512***	0.000509***
		(0.000141)	(0.000138)	(0.000142)	(0.000139)	(0.000040)	(0.000041)	(0.000041)
	Pupil-teacher ratio	-0.035725**	-0.036461**	-0.036511**	-0.037783**	-0.071267***	-0.070629***	-0.071777***
		(0.014506)	(0.014439)	(0.014976)	(0.014759)	(0.004851)	(0.004975)	(0.004960)
	% free/reduced price lunch	0.002839	0.002548	0.003981	0.003309	-0.001184	0.000122	-0.000488
		(0.004770)	(0.004831)	(0.004846)	(0.004842)	(0.001743)	(0.001821)	(0.001818)
	% ELL	-0.001326	-0.001434	-0.001077	-0.001221	-0.002647***	-0.002450**	-0.002436**
		(0.003212)	(0.003212)	(0.003201)	(0.003201)	(0.000963)	(0.000966)	(0.000965)
	% SPED	0.005713*	0.005928*	0.005884*	0.006032*	0.008713***	0.008634***	0.008722***
		(0.003238)	(0.003183)	(0.003244)	(0.003169)	(0.001289)	(0.001299)	(0.001297)

	Pupil-teacher aide ratio	0.000358	0.000376	0.000381	0.000393	0.000512***	0.000491***	0.000524***
		(0.000367)	(0.000370)	(0.000366)	(0.000370)	(0.000117)	(0.000117)	(0.000118)
Teacher characteristics	Age	-0.006826***	-0.007474***	-0.006719***	-0.007297***			
		(0.001976)	(0.001917)	(0.001951)	(0.001884)			
	Female	-0.221774***	-0.241228***	-0.227531***	-0.244677***			
		(0.042192)	(0.042793)	(0.042952)	(0.042834)			
	Teacher of color	-0.113029***	-0.113643***	-0.061841	-0.061686			
		(0.039723)	(0.039849)	(0.039574)	(0.039749)			
	Advanced degree	-0.028580	-0.037085	-0.020382	-0.031736			
		(0.033897)	(0.033126)	(0.035510)	(0.033980)			
	Mid-career (4-10 years)	-0.336968***	-0.333981***	-0.341111***	-0.335867***			
		(0.057962)	(0.056872)	(0.058044)	(0.056569)			
	Late career (over 10 years)	-0.378946***	-0.370086***	-0.371535***	-0.366649***			
		(0.053621)	(0.053066)	(0.053419)	(0.052525)			
	TE: Value-added (z-score)	-0.014946		-0.018445				
		(0.040416)		(0.039768)				
	TE: Student survey (z-score)	-0.032681		-0.040182				
		(0.026285)		(0.026764)				
	TE: SOEI (z-score)	-0.163376***		-0.171417***				
		(0.031534)		(0.033825)				
	TE: Composite (z-score)		-0.188266***		-0.204352***			
			(0.035822)		(0.036721)			
eractions between her characteristics	TE: SOEI X Pupil-teacher ratio			-0.002475			0.004304	
				(0.009123)			(0.005536)	
	TE: SOEI X Avg. Bonus			0.000166*			0.000193**	
				(0.000087)			(0.000075)	
	TE: SOEI X			-0.000973			0.000207	

% proficient in reading			(0.001629)			(0.001037)	
TE: SOEI X % students of color			-0.005826***			-0.009427***	
			(0.001398)			(0.001175)	
Teacher of color X % students of color			0.009998***	0.010151***		0.011047***	0.011626***
			(0.001603)	(0.001557)		(0.001579)	(0.001572)
TE: Composite X Pupil-teacher ratio				0.000081			0.000192**
				(0.000088)			(0.000087)
TE: Composite X Avg. Bonus				-0.006167			-0.002913
				(0.010365)			(0.006715)
TE: Composite X % proficient in reading				0.000873			0.001588
				(0.001554)			(0.001189)
TE: Composite X % students of color				-0.003864***			-0.007268***
				(0.001434)			(0.001293)
Year FE	Y	Y	Y	Y	Y	Y	Y
Teacher FE	N	N	N	N	Y	Y	Y
Log-likelihood	-165644.4	-165913.0	-165168.9	-165582.7	-130671.9	-129184.0	-129436.0
Chi2	27414.9	26546.8	35819.3	27888.7	2366.5	6922.2	6628.9
Observations	1,766,858	1,766,858	1,766,858	1,766,858	1,957,233	1,957,233	1,957,233

Note: The dependent variable is a binary choice for teacher *i* whether to apply to a position at school *j* in year *y*. For models T1A-T2B, we use the logistic regression to estimate teacher preferences and the standard errors are clustered at the sending-school level. For models T3-T5, we use the conditional mixed logistic regression and the standard errors are clustered at the teacher level. The results are based on a subsample of all active MPS teachers who meet the licensure requirement for each position posted in I&S in each year. As such, teachers may have varying choice sets available to them. For teachers who do not participate in I&S, we assume that they choose their own school out of all other alternatives available to them. We also pool in information from the post-interview stage for teachers who received multiple job offers. Indicators for whether the observation is from the post-offer stage and from which round are included in all models. Estimates are reported in the log-odds format. *p < .10. **p < 0.05. ***p < .01.

Table 2.3: Marginal Rates of Substitution Matrix

	Bonus (dollars)	Commute Time (minutes)	High-priority school status	Avg. teacher experience (year)	% students of color	% teachers of color	Reading achievement (% proficient)	Enrollment	Pupil-teacher ratio
Bonus (dollars)	-1	0.01	0	-0.03	0.14	0.25	-0.14	-1.87	0.01
Commute Time (minutes)	84.3	-1	-0.25	2.68	-11.9	-21.4	11.44	158	-1.15
High-priority school status	332	-3.94	-1	10.6	-47	-84.3	45	621	-4.52
Avg. teacher experience (year)	-31.4	0.37	0.09	-1	4.45	7.98	-4.26	-58.8	0.43
% students of color	7.06	-0.08	-0.02	0.22	-1	-1.79	0.96	13.2	-0.1
% teachers of color	3.94	-0.05	-0.01	0.13	-0.56	-1	0.53	7.37	-0.05
Reading achievement (% proficient)	-7.37	0.09	0.02	-0.23	1.04	1.87	-1	-13.8	0.1
Enrollment	-0.53	0.01	0	-0.02	0.08	0.14	-0.07	-1	0.01
Pupil-teacher ratio	73.5	-0.87	-0.22	2.34	-10.41	-18.66	9.97	138	-1

Note: For each cell, we use the estimated preferences from column T3 of Table 1 to calculate the rate at which a teacher is ready to face a one unit increase in the column attribute in exchange for the row attribute while maintaining the same level of utility. Negative values indicate that the teacher must forgo the column attribute. For example, a high priority school status requires a reduction of 47 percentage points in the share of students of color to keep the predicted likelihood that an average teacher applies to that school constant. Although the magnitudes of the estimates shown above can be impractically large, a change in any attribute can be fully or partially neutralized by changing multiple attributes at the same time. Doing so gives the district the option to avoid changing non-malleable attributes or making an extremely large change in one single attribute. To neutralize the effect of a change in high priority school status, the district may decrease the number of students per teacher at that school by roughly 2.25 students and provide any applicant to that school a bonus of 116 dollars. Alternatively, the district can also just provide any applicant to that school a bonus of 332 dollars without altering the pupil-teacher ratio.

Table 2.4: Characteristics of schools in the bottom and top quintiles of student reading proficiency

School Attributes	Mean at Schools in the Bottom Quintile	Mean at Schools in the Top Quintile	Difference
% students proficient in reading	22.5	81.8	-59.3
% priority schools	0.24	0	24
Enrollment	437	830	-393
Avg. teacher experience, yrs	10.9	17.4	-6.50
% teachers of color	17.8	7.8	10
% students of color	91.4	33.9	57.5
% low-income students	89.0	27.9	61.1
% ELL	25.9	7.44	18.5
% Special education	32.8	11.3	21.5
Pupil-teacher ratio	11.1	19.9	-8.80
Pupil-teacher aide ratio	25.9	80.1	-54.2

Table 2.5: Change in bottom-quintile school characteristics required to create indifference between working in top-quintile and bottom-quintile school

Type of Teacher	Change Required at Bottom Quintile School				Cost to Change (per teacher)
	Pay (method 1)	Pay (method 2)	Pupil-Teacher Ratio	Pupil-Aide Ratio	Pupil-Teacher Ratio
Average teacher	$746 (354,1264)	$658 (387,880)	-17.4 (-19.9,-14.0)	1,142 (385,3106)	$45,496
Effectiveness 1SD below average	$67.0 (-495,810)	$47.8 (-579,561)	-9.38 (-15.9,-0.87)	32.8 (-711,1960)	$42,252
Effectiveness 1SD above average	$1,424 (1203,1717)	$1,269 (1199,1354)	-25.4 (-27.2,-24.0)	2,253 (1481,4253)	$45,496
Teacher of color	-$36.1 (-514,597)	-$45.0 (-602,410)	-8.16 (-14.0,-0.57)	-136 (-736,1419)	$36,757

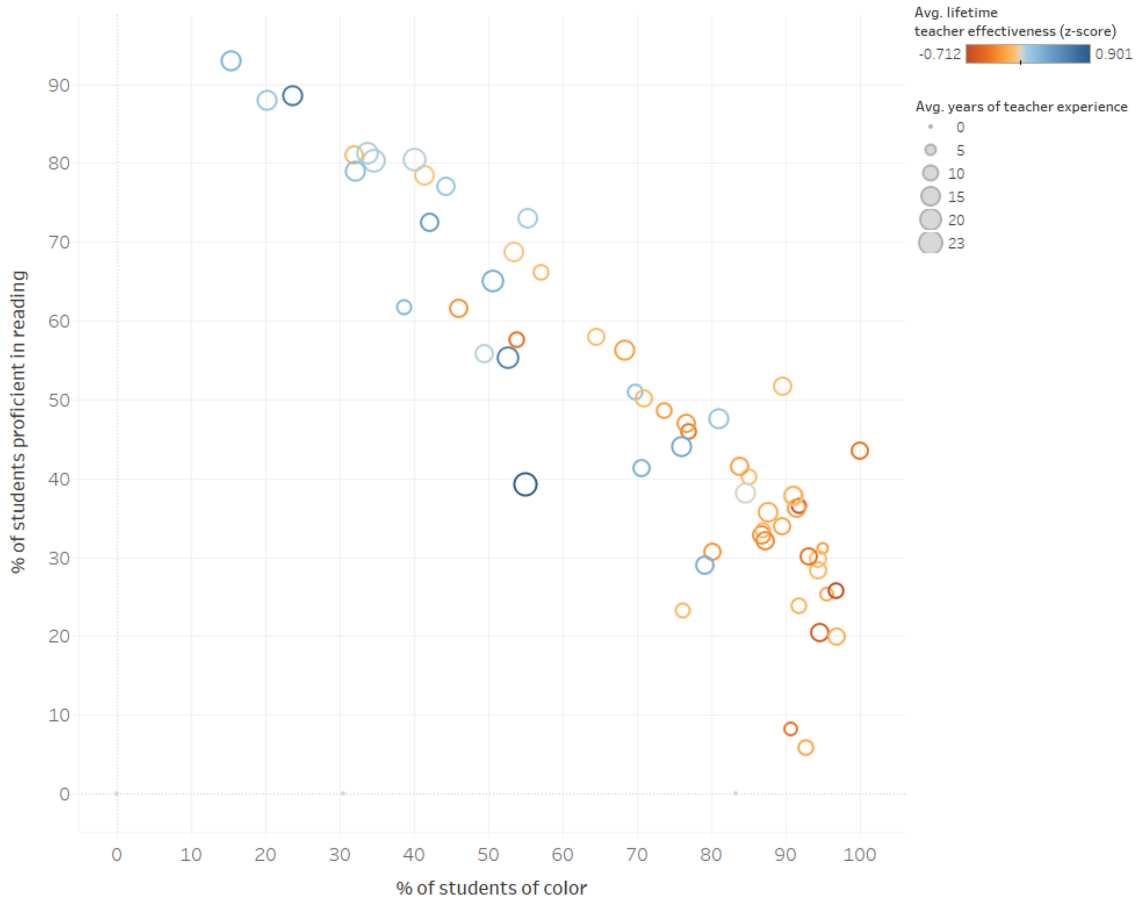
Note: To obtain the requisite changes, we divide the coefficients of school attributes from model T4B by the coefficient on each lever to obtain the marginal rates of substitution. Then, we multiply the vector of attribute differences shown in Table 4 with the corresponding marginal rates of substitution to obtain the requisite change in each lever for each attribute. For each lever, summing up all the requisite changes across all attributes yields the total requisite change to create indifference for an average teacher. For specific types of teachers, we obtain the differential change from the coefficients of the interaction terms. Method 1 uses pay-for-performance variation. Method 2 uses estimated cost of travel time ($22.9 per hour) that is established in the transportation literature and assumes 200 work days in a year. Cost to change pupil-teacher ratio is calculated using the average enrollment at bottom-quintile schools and assume that the total cost of hiring an additional teacher is $50,000 per year. When the resulting pupil-teacher ratio drops below zero, we use a ratio of one instead. Confidence intervals are included in parentheses.

Table 2.6: School Preferences Results

	(S1A)	(S1B)	(S2A)	(S2B)
Age	-0.020693***	-0.020704***	-0.020690***	-0.020700***
	(0.002759)	(0.002747)	(0.002755)	(0.002744)
Female	0.104905**	0.109752**	0.104691**	0.109528**
	(0.051508)	(0.050946)	(0.051576)	(0.051025)
Teacher of color	-0.247915***	-0.246061***	-0.219973	-0.215099
	(0.059363)	(0.059392)	(0.154204)	(0.156329)
Holding an advanced degree	0.158620***	0.154168***	0.158630***	0.154174***
	(0.057105)	(0.056724)	(0.057103)	(0.056726)
Mid-career (4-10 years)	0.053051	0.060893	0.053122	0.060966
	(0.069116)	(0.069605)	(0.068973)	(0.069457)
Late-career (over 10 years)	0.148065*	0.149136*	0.148398*	0.149485*
	(0.081075)	(0.081153)	(0.081000)	(0.081056)
Early offers candidate	0.091804	0.081764	0.091978	0.081960
	(0.149591)	(0.151511)	(0.149747)	(0.151662)
Automatically interview	1.296517***	1.296620***	1.296495***	1.296598***
	(0.053095)	(0.053104)	(0.053090)	(0.053101)
Average number of years at all previous schools	0.072413***	0.078991***	0.072336***	0.078901***
	(0.018389)	(0.017728)	(0.018357)	(0.017689)
TE: SOEI (z-score)	0.243779***		0.243753***	
	(0.058059)		(0.058038)	
TE: Student survey (z-score)	0.110648*		0.110824*	
	(0.057088)		(0.056895)	
TE: Value-added (z-score)	0.053806		0.053626	
	(0.066413)		(0.066179)	
TE: Composite (z-score)		0.289941***		0.289990***
		(0.055820)		(0.055839)
Teacher of color X % student of color at receiving school			-0.000386	-0.000428
			(0.001960)	(0.001998)
Log-likelihood	-5691.0	-5695.9	-5691.0	-5695.9
Chi2	2110.1	1763.3	2115.5	1765.8
Observations	12,427	12,427	12,427	12,427

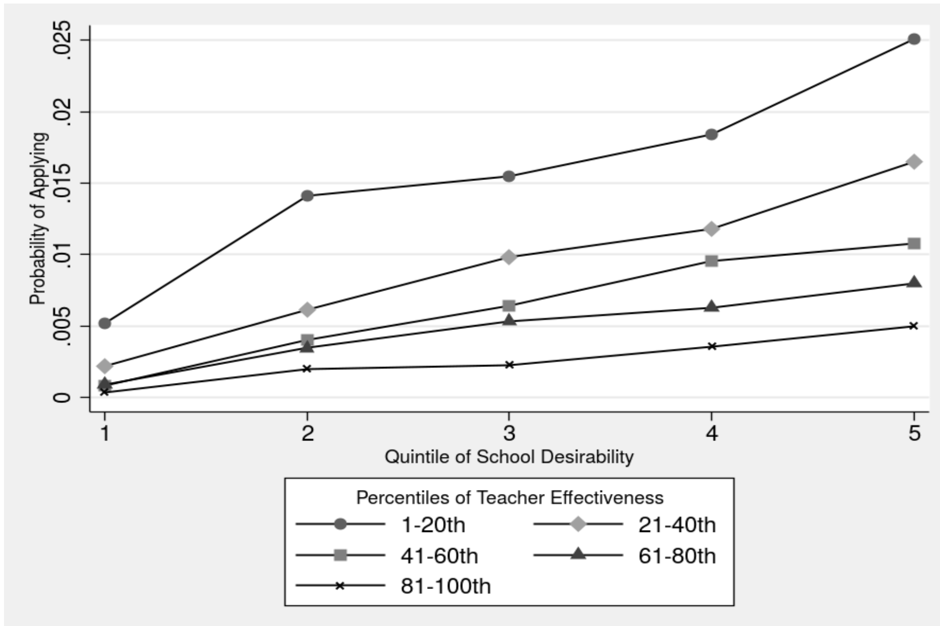
Note: The above results are obtained from running rank-ordered logistic regressions where the receiving-schools are the decision makers. The choice sets are the pools of applications. For all models, the dependent variable is an ordered ranking from 1 to 4, where 1 is the highest ranking. Applicants who were screened out before the interview stage receive a censored ranking of zero. We omit automatic interview applicants who were not ranked in the top four because we cannot observe whether they were revealed preferred to other candidates who were not interviewed. The difference between A and B variants of these models is we break down teacher effectiveness measures into the three sub-components in A and combine them in B. The effectiveness measures are the most recent rolling averages from the first year of work up to the year of application. For more information on the teacher effectiveness measures, see Appendix B. Robust standard errors are clustered at the school level. *p < .10. **p < 0.05. ***p < .01.

Figure 2.1: Average School Characteristics at MPS (2009-2015)

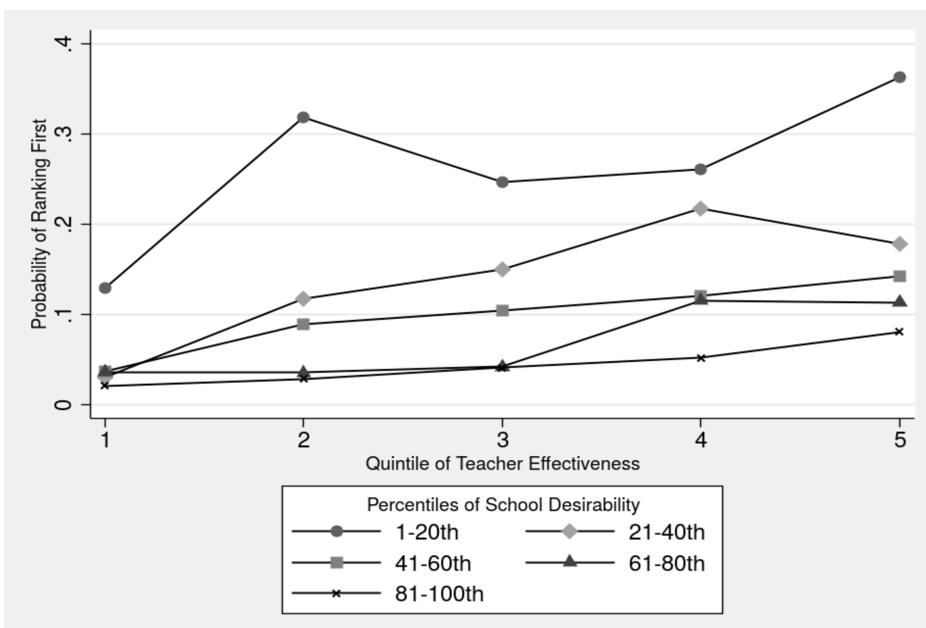


Note: The characteristics displayed above are the averages at the school level between 2009 and 2015. Teacher effectiveness is measured by a score composite that takes into account every type of available measure of effectiveness.

Figure 2.2: Evidence on Strategic Behavior

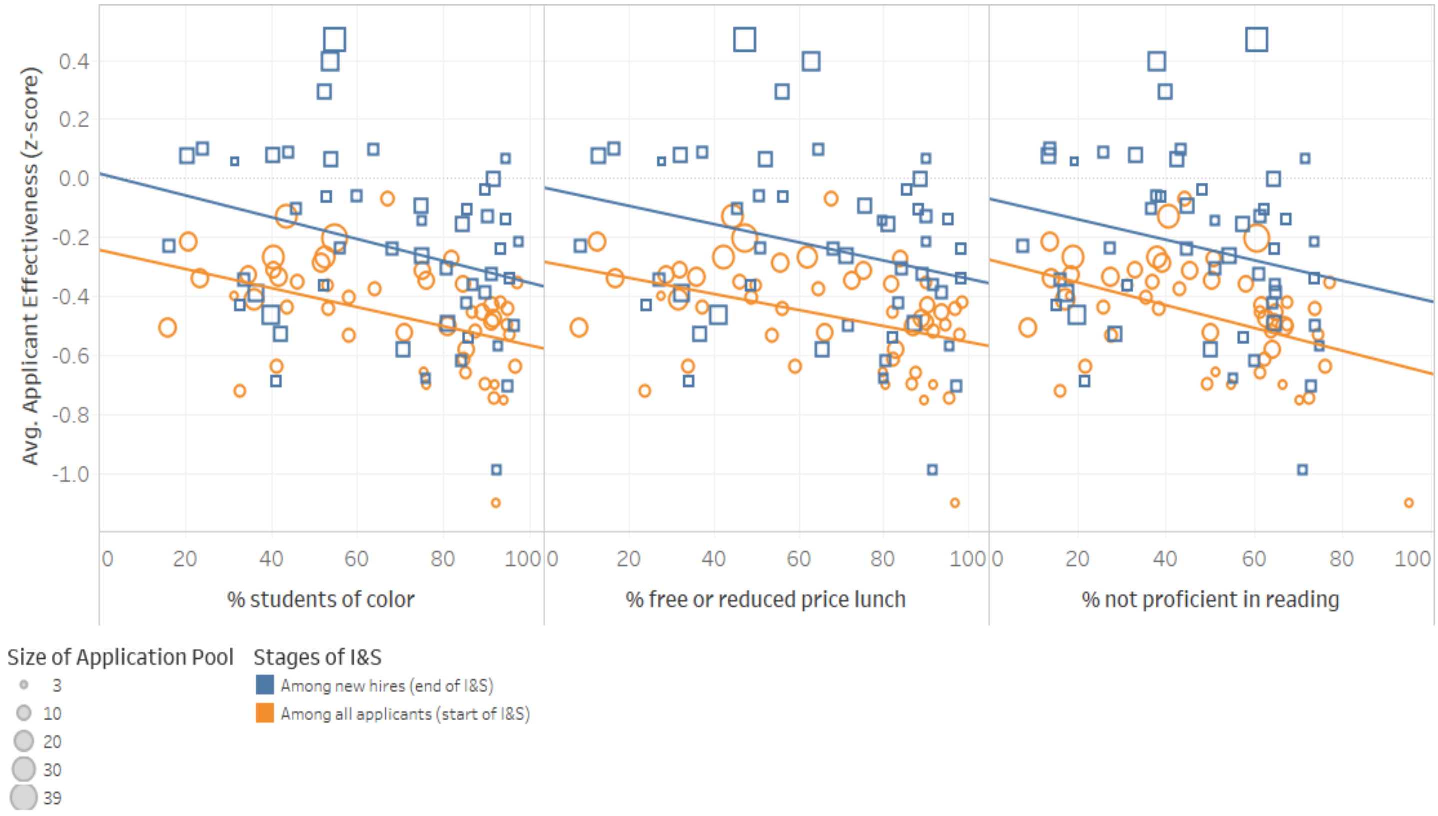


Panel A: Test of Strategic Behavior among Teachers



Panel B: Panel A: Test of Strategic Behavior among Schools

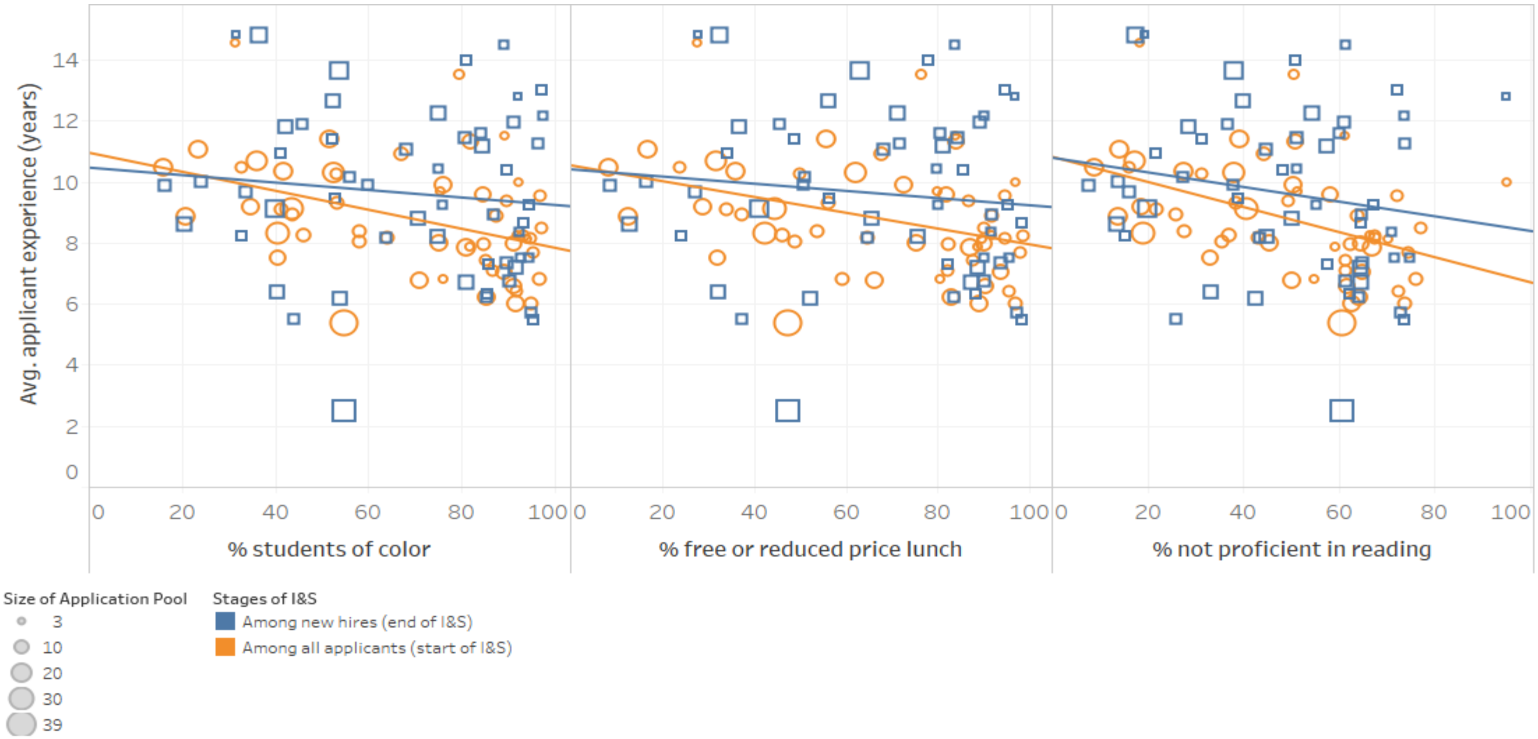
Figure 2.3: Average Applicant Effectiveness as a Function of Receiving-school Characteristics by Stage of I&S (2013-2015)



Note: Average receiving-school characteristics are on the x-axes. Average applicant effectiveness among every teacher who applied to (orange) and those who were hired (blue) at each school at the end of the I&S process is on the y-axis. Teacher effectiveness is measured by a running average of the score composite that takes into account every type of available measure of effectiveness for each teacher from her first year at MPS up to the year of application.

Figure 2.4: Average Applicant Experience as a Function of Receiving-school Characteristics

by Stages of I&S (2009-2015)



Note: Average receiving-school characteristics are on the x-axes. Average applicant experience among every teacher who applied to (orange) and those who were hired (blue) at each school at the end of the I&S process is on the y-axis.

Figure 2.5: Average Size of the Application Pool as a Function of Receiving-school Characteristics (2009-2015)

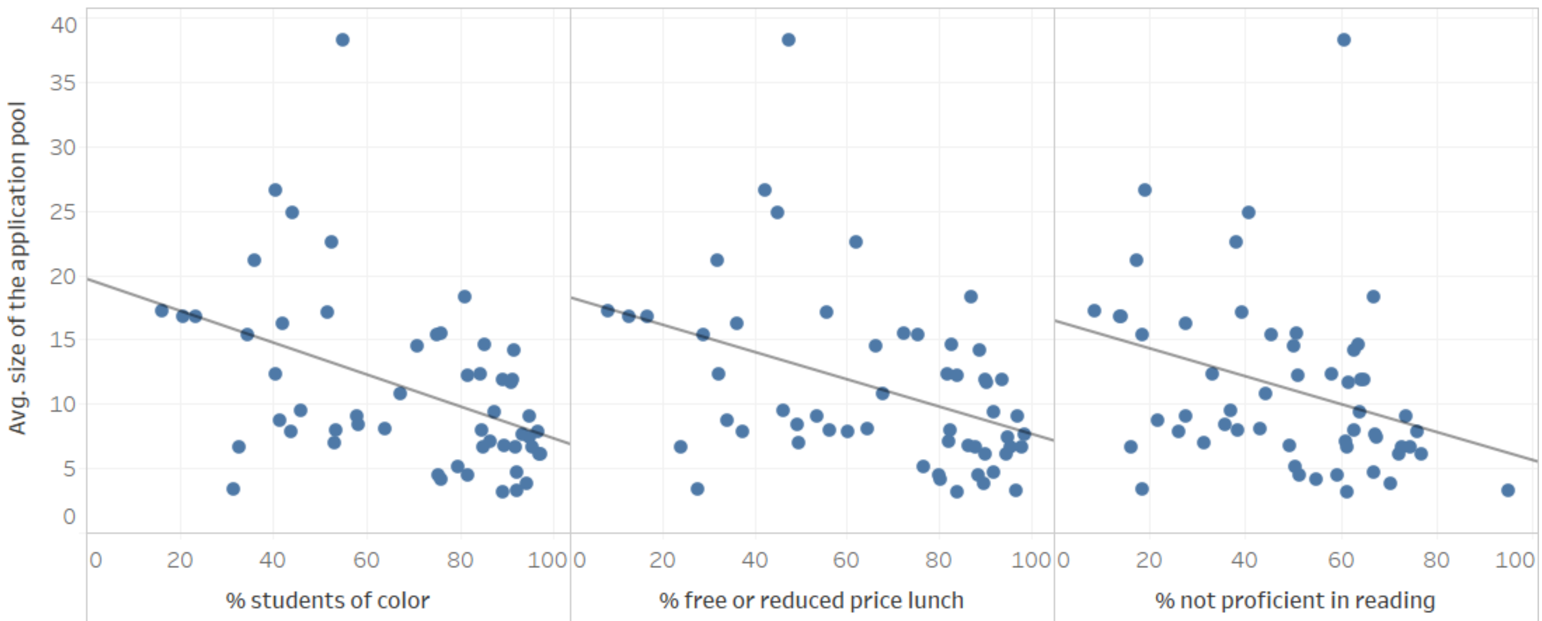


Table 3.1: Summary Statistics of Test Takers within 5 Miles from Major Airports in Texas (2003-2008)

Variable	DFW		Austin		IAH		Hobby		San Antonio		Texas ex.DFW	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
# test takers	136	136	93	100	235	209	**136**	**148**	**120**	**136**	**114**	**118**
	118	119	95	105	193	219	143	163	136	144	130	134
Fraction black	**0.13**	**0.15**	0.20	0.18	**0.31**	**0.39**	0.29	0.29	0.12	0.11	0.14	0.14
	0.13	0.12	0.17	0.14	0.13	0.16	0.33	0.32	0.16	0.14	0.21	0.20
Fraction Hispanic	**0.21**	**0.26**	**0.73**	**0.76**	0.41	0.43	0.63	0.65	**0.55**	**0.59**	**0.40**	**0.43**
	0.14	0.18	0.17	0.15	0.12	0.14	0.32	0.31	0.25	0.25	0.32	0.32
Fraction white	**0.55**	**0.47**	**0.066**	**0.050**	**0.25**	**0.15**	**0.057**	**0.037**	**0.31**	**0.27**	**0.43**	**0.40**
	0.24	0.25	0.071	0.056	0.19	0.15	0.081	0.053	0.23	0.22	0.33	0.32
Fraction econ. disadvantaged	**0.31**	**0.39**	**0.81**	**0.85**	**0.61**	**0.67**	**0.80**	**0.82**	**0.56**	**0.60**	**0.54**	**0.56**
	0.23	0.26	0.15	0.13	0.18	0.20	0.18	0.14	0.29	0.29	0.28	0.27
Fraction ESL	**0.059**	**0.070**	**0.061**	**0.072**	0.073	0.083	0.046	0.042	**0.013**	**0.020**	**0.043**	**0.047**
	0.061	0.063	0.070	0.094	0.067	0.069	0.070	0.063	0.031	0.033	0.083	0.083
Fraction gifted	0.14	0.14	**0.066**	**0.057**	**0.082**	**0.054**	**0.060**	**0.073**	**0.067**	**0.082**	**0.10**	**0.094**
	0.010	0.080	0.052	0.046	0.048	0.048	0.064	0.069	0.072	0.073	0.10	0.099
Fraction female	**0.50**	**0.50**	**0.51**	**0.49**	0.49	0.49	0.51	0.51	0.50	0.50	**0.50**	**0.50**
	0.055	0.068	0.071	0.074	0.055	0.062	0.073	0.069	0.096	0.074	0.10	0.088
Fraction male	**0.50**	**0.50**	**0.49**	**0.51**	0.51	0.51	0.49	0.49	0.50	0.50	**0.50**	**0.50**
	0.055	0.068	0.071	0.074	0.055	0.063	0.073	0.068	0.096	0.074	0.10	0.088
Scores	**2255**	**2301**	**2147**	**2181**	2178	2193	**2166**	**2208**	**2194**	**2254**	**2204**	**2254**
	93	100	82	90	77	99	93	85	111	102	99	97
N	440	456	504	524	200	311	1172	1100	1110	1075	117732	122492

Note: Post-closure is defined as being after January 1st, 2006. Mean values are presented on the top row of each variable with standard deviation underneath. The means that are statistically different at 5% significance level between pre and post periods are bolded.

Table 3.2: Difference-in-Differences Results by Outcome and Comparison Group

	Austin	IAH	Hobby	San Antonio	Airports Aggregate	Texas ex.DFW
	(1)	(2)	(3)	(4)	(5)	(6)
Scores	0.028***	0.01	0.023**	0.020**	0.023***	0.016*
	(0.009)	(0.009)	(0.01)	(0.009)	(0.009)	(0.009)
Passing rate	0.712	-0.061	-0.168	0.596	0.389	-0.007
	(0.953)	(0.8)	(0.775)	(0.757)	(0.576)	(0.458)
Commended performance rate	7.271***	4.110***	6.585***	3.665***	5.551***	3.969***
	(1.356)	(1.404)	(1.666)	(1.387)	(1.485)	(1.485)
Absence rate	0.100	-0.497**	-0.347**	-0.002	-0.172	-0.122
	(0.174)	(0.237)	(0.153)	(0.135)	(0.119)	(0.135)
N	1,924	1,407	3,178	3,079	9,588	240,987

Note: All models include day-of-week, year-grade, school-grade, and month fixed effects. Other controls include characteristics of the composition of test takers on each test date. Scores are in z-values while the rest of the outcomes are in percentage. In columns 1-5, observations with larger number of test takers and with location closer to airports receive more weight. Observations with larger number of test takers receive more weight in column 6. Standard errors are clustered at the campus level in parentheses. *p < .10. **p < 0.05. ***p < .01.

Table 3.3: Results by Outcome, Comparison Group, and Level of Education

		Austin	IAH	Hobby	San Antonio	Airports Aggregate	Texas ex.DFW
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High school	Scores	0.028***	0.036***	0.047***	0.044***	0.045***	0.038***
		(0.006)	(0.011)	(0.009)	(0.009)	(0.007)	(0.007)
	Passing rate	-0.379	1.65	1.878	2.615	2.195	1.965
		(1.22)	(2.031)	(2.434)	(2.253)	(1.928)	(1.645)
	Commended performance rate	7.743***	8.331***	9.432***	7.988***	8.694***	6.909***
		(1.176)	(0.692)	(1.611)	(1.267)	(1.486)	(2.168)
	N	240	193	400	436	1,269	37,452
Panel B: Elementary school	Scores	0.033***	0.022*	0.012	0.023***	0.022***	0.012
		(0.011)	(0.011)	(0.011)	(0.009)	(0.008)	(0.008)
	Passing rate	5.208***	3.012**	0.892	1.763**	2.287***	0.800
		(1.303)	(1.175)	(1.069)	(0.701)	(0.75)	(0.626)
	Commended performance rate	3.838***	3.541***	3.398***	2.943**	3.496***	2.281**
		(1.112)	(1.28)	(1.216)	(1.213)	(1.051)	(1.024)
	N	830	612	1,446	1,257	4,145	91,839

Note: Observations from grade 9 to 10 and from grade 3 to 4 are included in Panel A and Panel B, respectively. All models include day-of-week, year-grade, school-grade, and month fixed effects. Other controls include characteristics of the composition of test takers on each test date. Scores are in z-values while the rest of the outcomes are in percentage. In columns 1-5, observations with larger number of test takers and with location closer to airports receive more weight. Observations with larger number of test takers receive more weight in column 6. Standard errors are clustered at the campus level in parentheses. *p < .10. **p < 0.05. ***p < .01.

Table 3.4: Robustness Check with Unit-specific Time Trends

	Austin	IAH	Hobby	San Antonio	Airports Aggregate	Texas ex.DFW
	(1)	(2)	(3)	(4)	(5)	(6)
Scores	0.029***	0.018*	0.021**	0.011	0.019**	0.016*
	(0.0090)	(0.010)	(0.010)	(0.0090)	(0.008)	(0.0090)
Passing Rate	1.137	0.204	0.924	0.523	0.877	-0.016
	(1.124)	(1.337)	(1.074)	(1.074)	(0.929)	(0.457)
Commended performance rate	3.189***	2.216*	2.346**	1.651	2.314**	3.982***
	(1.208)	(1.196)	(1.128)	(1.118)	(1.03)	(1.485)
N	1,924	1,407	3,178	3,079	9,588	240,987

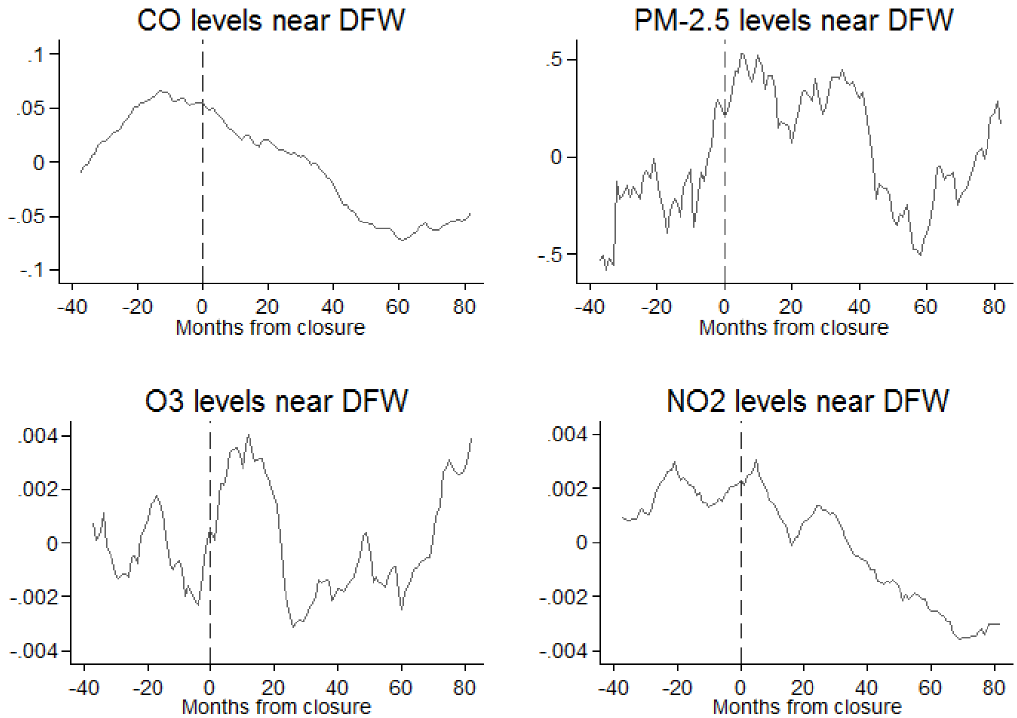
Note: The difference between the models in this table and the models used to generate Table 2 is the inclusion of unit-specific time trends. I include an interaction term between time trend and being near DFW. All models include day-of-week, year-grade, school-grade, and month fixed effects. Other controls include characteristics of the composition of test takers on each test date. Scores are in z-values while the rest of the outcomes are in percentage. In columns 1-5, observations with larger number of test takers and with location closer to airports receive more weight. Observations with larger number of test takers receive more weight in column 6. Standard errors are clustered at the campus level in parentheses. *p < .10. **p < 0.05. ***p < .01.

Table 3.5: Robustness Check with Placebo Treatment

	Austin	IAH	Hobby	San Antonio	Airports Aggregate	Texas ex.DFW
	(1)	(2)	(3)	(4)	(5)	(6)
Scores	-0.017**	-0.008	-0.01	0.003	-0.006	-0.001
	(0.008)	(0.01)	(0.007)	(0.007)	(0.007)	(0.01)
Passing Rate	-0.917	0.654	-0.163	-0.041	-0.324	0.403
	(0.961)	(1.249)	(0.975)	(0.686)	(0.682)	(0.838)
Commended performance rate	1.384	1.235	0.44	1.013	0.781	0.469
	(0.91)	(1.083)	(0.856)	(0.89)	(0.772)	(1.023)
N	626	420	1,080	1,028	3,154	78,247

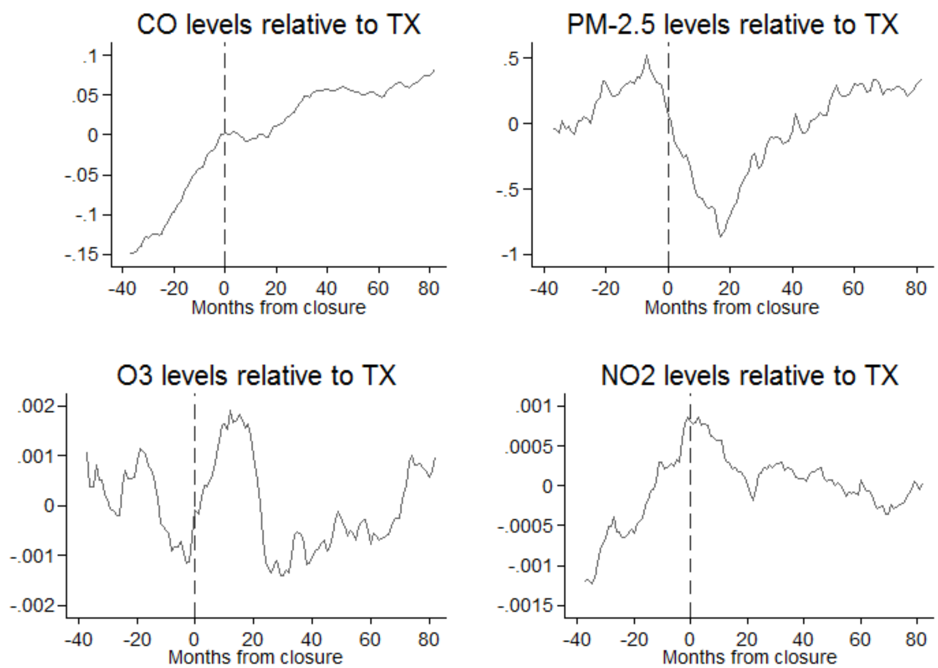
Note: Treatment year in these regressions is redefined as 2004 instead of 2006. Unrelated outcome variable is randomly generated. All models include day-of-week, year-grade, school-grade, and month fixed effects. Other controls include characteristics of the composition of test takers on each test date. Scores are in z-values while the rest of the outcomes are in percentage. In columns 1-5, observations with larger number of test takers and with location closer to airports receive more weight. Observations with larger number of test takers receive more weight in column 6. Standard errors are clustered at the campus level in parentheses. *p < .10. **p < 0.05. ***p < .01.

Figure 3.1: Monthly Pollution Averages around DFW by Pollutant



Note: Monthly pollution averages from the closest monitor to the airport are regressed on month fixed effects to reduce the effects of seasonality. The residuals are plotted over time.

Figure 3.2: Monthly Pollution Averages around DFW Relative *to the Rest of Texas* by Pollutant



Note: Monthly pollution averages from the closest monitor to the airport are regressed on month fixed effects to reduce the effects of seasonality. The residuals are plotted over time.

Bibliography

Airports Council International. 2005. "Passenger Traffic 2005 Final." Retrieved from http://www.aci.aero/Data-Centre/Annual-Traffic-Data/Passengers/2005-final.

Ahn, Tom. 2015. "Matching Strategies of Teachers and Schools in General Equilibrium." *IZA Journal of Labor Economics* 4, no. 1.

Aldy, J. and A. M. Bind. 2013. "Life-saving Information: Mortality Risk Reduction from Air Quality Index Forecasts." Harvard University, Working Paper.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *The Quarterly Journal of Economics*. 114, no.4: 1243-1284.

Alesina, Alberto, and Ferrara, Eliana La. 2000. "Participation in Heterogeneous Communities." The *Quarterly Journal of Economics* 115, no. 3: 847-904.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003."Fractionalization." *Journal of Economic growth* 8, no. 2: 155-194.

Allison, Paul D., and Richard P. Waterman. 2002. "Fixed–Effects Negative Binomial Regression Models." *Sociological methodology* 32, no. 1: 247-265.

Allouch, Nizar. "On the Private Provision of Public Goods on Networks." 2015. *Journal of Economic Theory* 157: 527-552.

Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo. 2008. "Inequality and Public Good Provision: An Experimental Analysis." *The Journal of Socio-Economics* 37, no. 3: 1010-1028.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100, no. 401: 464-477.

Andreoni, James, and A. Abigail Payne. 2013. "Charitable Giving." *Handbook of Public Economics* 5: 1-50.

Barnes, Gary, and Peter Langworthy. 2003. "The per-mile costs of operating automobiles and trucks." Minnesota Department of Transportation.

Becker,Gary. 1957. *The Economics of Discrimination*. Chicago. University of Chicago Press.

Beggs, Steven, Scott Cardell, and Jerry Hausman. 1981. "Assessing the potential demand for electric cars." *Journal of Econometrics* 17, no. 1: 1-19.

Belenky, Peter. 2011. "Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis." U.S. Department of Transportation.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian.1986. "On the Private Provision of Public Goods." Journal of Public Economics 29, no. 1: 25-49.

Béteille, Tara, Demetra Kalogrides, and Susanna Loeb. 2012. "Stepping stones: Principal career paths and school outcomes." *Social Science Research* 41, no. 4: 904-919.

Bonhomme, Stéphane, Grégory Jolivet, and Edwin Leuven. 2016. "School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities." *The Economic Journal* 126: 1342-1371.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students." *The American Economic Review* 95, no. 2: 166-171.

Boyd, Donald, Hamp Lankford, Susanna Loeb, Matthew Ronfeldt, and Jim Wyckoff. 2011. "The Role of Teacher Quality in Retention and Hiring: Using Applications to Transfer to Uncover Preferences of Teachers and Schools." *Journal of Policy Analysis and Management* 30, no. 1: 88-110.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2013. "Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers." *Journal of Labor Economics* 31, no. 1: 83-117.

Branch, Gregory, Eric Hanushek, and Steven Rivkin. 2009. "Estimating Principal Effectiveness." National Center for Analysis of Longitudinal Data in Education Research Working Paper no. 32.

Burbidge, John B., Lonnie Magee and A. Leslie Robb. 1988 "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical*

Association, vol. 83, 123-127.

Bureau of Labor Statistics. 2005. “Economy at a Glance.” Retrieved from https://www.bls.gov/eag/eag.tx_dallas_msa.htm.

Bussing, Regina, Neal Halfon, Bernadette Benjamin, and Kenneth B. Wells. 1995. "Prevalence of behavior problems in US children with asthma." *Archives of Pediatrics & Adolescent Medicine* 149, no. 5: 565-572.

Caiazzo, Fabio, Akshay Ashok, Ian A. Waitz, Steve HL Yim, and Steven RH Barrett. 2013. "Air Pollution and Early Deaths in the United States. Part I: Quantifying the Impact of Major Sectors in 2005." *Atmospheric Environment* 79: 198-208.

Chan, Kenneth S., Stuart Mestelman, Robert Moir, and R. Andrew Muller. 1999. "Heterogeneity and the Voluntary Provision of Public Goods." *Experimental Economics* 2, no. 1: 5-30.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113, no. 2: 376-424.

Chapman, Randall G., and Richard Staelin. 1982. "Exploiting rank ordered choice set data within the stochastic utility model." *Journal of Marketing Research*: 288-301.

Chen, L., B. L. Jennison, W. Yang, and S. T. Omaye: 2000. “Elementary School Absenteeism and Air Pollution.” *Inhalation Toxicology* 12(11): 997– 1016.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics* 126, no. 4: 1593-1660.

Clotfelter, Charles T., Helen F. Ladd, and Jacob Vigdor. 2005. "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education Review* 24, no.4: 377-392.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41, no.4: 778-820.

Currie, Janet, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin. 2009. "Does Pollution Increase School Absences?" *The Review of Economics and Statistics* 91, no. 4: 682-694.

Currie, Janet, and Duncan Thomas. 2012. "Early Test Scores, School Quality and SES: Long run Effects on Wage and Employment Outcomes." *Research in Labor Economics* 35: 185-214.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy* 107, no. 3: 455-506.

Dallas/Fort Worth International Airport.: 2004, 'Questions and Answers Regarding Delta Airlines' Restructuring at DFW Airport: September 8, 2004'. Press Report, Dallas/Fort Worth International Airport.

Dallas/Fort Worth International Airport. 2006. "Comprehensive Annual Financial Report." Technical Report. Dallas/Fort Worth International Airport.

David, H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the US Labor Market." *American Economic Review* 96, no. 2: 189-194.

DeArmond, Michael, Betheny Gross, and Dan Goldhaber. 2010. "Is it Better to be Good or Lucky? Decentralized Teacher Selection in 10 Elementary Schools." *Educational Administration Quarterly* 46, no. 3: 322-362.

Deutsch, Karl. 1966. *Nationalism and Social Communication*. Cambridge: The MIT Press.

Ejermo, Olof. 2005. "Technological Diversity and Jacobs' Externality Hypothesis Revisited." *Growth and Change* 36, no. 2: 167-195.

Feigenbaum, James J., and Andrew B. Hall. 2015. "How High-Income Neighborhoods Receive More Service from Municipal Government: Evidence from City Administrative Data." http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum_hall_respublica.pdf

Feng, Li, and Tim R. Sass. 2016. "Teacher Quality and Teacher Mobility." *Education Finance and Policy* 1: 1-41.

Fowler, M. G., M. G. Davenport, and R. Garg: 1992. “School Functioning of US Children with Asthma.” *Pediatrics* 90(6): 939–944.

Frey, William H. 2014. *Diversity Explosion: How New Racial Demographics are Remaking America*. Brookings Institution Press.

Gale, David, and Lloyd S. Shapley. "College admissions and the stability of marriage." *The American Mathematical Monthly* 69, no. 1 (1962): 9-15.

Gates, Susan M., Jeanne S. Ringel, and Lucrecia Santibanez. 2003. *Who Is Leading Our Schools?: An Overview of School Administrators and Their Careers*. Rand Corporation.

Gauderman, W. James, et al. 2004. "The Effect of Air Pollution on Lung Development from 10 to 18 Years of Age." *New England Journal of Medicine* 351.11: 1057-1067.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities." *Journal of Political Economy* 100, no. 6: 1126-1152.

Gilliland, F. D., K. Berhane, E. B. Rappaport, D. C. Thomas, E. Avol, W. J. Gauderman, S. J. London, H. G. Margolis, R. McConnell, K. T. Is- lam, et al. 2001. ‘The Effects of Ambient Air Pollution on School Absenteeism due to Respiratory Illnesses’. *Epidemiology* 12 (1): 43–54.

Goldhaber, Dan, Betheny Gross, and Daniel Player. 2007. "Are Public Schools Really Losing Their Best? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce.” National Center for Analysis of Longitudinal Data in Education Research. Working Paper no. 12.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?." *American Political Science Review* 101, no. 4: 709-725.

Halterman, Jill S., Kelly M. Conn, Emma Forbes-Jones, Maria Fagnano, A. Dirk Hightower, and Peter G. Szilagyi. 2006. "Behavior Problems among Inner-city Children with Asthma: Findings from a Community-based Sample." *Pediatrics* 117, no. 2: 192-199.

Ham, John C., Jacqueline S. Zweig, and Edward Avol. 2014. "Pollution, Test Scores and Distribution of Academic Achievement: Evidence from California Schools 2002–2008."

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Why Public Schools Lose Teachers." *Journal of Human Resources* 39, no. 2: 326-354.

Hanushek, Eric Alan, and Steven G. Rivkin. 2007. "Pay, Working Conditions, and Teacher Quality." *The Future of Children* 17, no.1: 69-86.

Hardin, Russell. 1995. *One for All: The Logic of Group Conflict*. Princeton: Princeton University Press.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship."*Econometrica* 52, no. 4: 909-38.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality In The United States, 1967–2006." *Review of Economic dynamics* 13, no. 1: 15-51.

Hitsch, Günter J., Ali Hortaçsu, and Dan Ariely. 2010. "Matching and Sorting in Online Dating." *The American Economic Review* 100, no. 1: 130-163.

Hsieh, Hsun-Ping, Tzu-Chi Yen, and Cheng-Te Li. 2015. "What Makes New York So Noisy?: Reasoning Noise Pollution by Mining Multimodal Geo-Social Big Data." *Proceedings of the 23rd ACM international conference on Multimedia*: 181-184.

Ingersoll, Richard M. 2003. "Out-of-Field Teaching and the Limits of Teacher Policy." *Center for the Study of Teaching and Policy*.

Isenberg, Eric, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, And Michael Hansen. 2013. "Access to Effective Teaching for Disadvantaged Students. NCEE 2014-4001." US Department of Education.

Jacobs, Jane. 1969. *The Economy of Cities*. New York: Vintage.

John B. Burbidge, Lonnie Magee, and A. Leslie Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable" *Journal of the American Statistical Association* 83, no. 401: 123-127.

Johnson, Scott D., and Christopher F. Roellke. 1999. "Secondary Teachers' and Undergraduate Education Faculty Members' Perceptions of Teaching-Effectiveness Criteria: A National Survey." *Communication Education* 48, no. 2: 127-138.

Klatte, Maria, Kirstin Bergström, and Thomas Lachmann. 2013. "Does Noise Affect Learning? A Short Review on Noise Effects on Cognitive Performance in Children." *Frontiers in Psychology*, 4.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." No. w23180. National Bureau of Economic Research.

Kopelman, Jason L., and Harvey S. Rosen. 2016. "Are Public Sector Jobs Recession-Proof? Were They Ever?" *Public Finance Review* 44, no. 3: 370-396.

Kremer, Michael. 1997. "How Much does Sorting Increase Inequality?" *The Quarterly Journal of Economics* 112, no. 1: 115-139.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24, no.1: 37-62.

Latane, Bibb, and John M. Darley. 1968. "Group Inhibition of Bystander Intervention in Emergencies." *Journal of Personality and Social Psychology* 10, no. 3: 215.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. 2014. "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation." No. w20648. National Bureau of Economic Research.

Legewie, Joscha., & Schaeffer, Merlin. 2016. "Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict." *American Journal of Sociology* 122, no.1: 125-161.

Levine, Jeremy R., and Carl Gershenson. 2014. "From political to material inequality: Race, immigration, and requests for public goods." *Sociological Forum* 29, no 3.

Loeb, Susanna, Demetra Kalogrides, and Eileen Lai Horng. 2010. "Principal Preferences and the

Uneven Distribution of Principals Across Schools." *Educational Evaluation and Policy Analysis* 32, no. 2: 205-229.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka (ed.), Frontiers in Econometrics, Academic Press, New York.

McFadden, D. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics*, 3, 303–328.

Miguel, Edward A. "Ethnic diversity and school funding in Kenya." *Center for International and Development Economics Research* (2001).

Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89, no.11: 2325-2368.

Minkoff, Scott L. 2016. "NYC 311 A Tract-Level Analysis of Citizen–Government Contacting in New York City." *Urban Affairs Review* 52, no 2: 211-246.

Montalvo, Jose G., and Marta Reynal-Querol. 2010. "Ethnic polarization and the duration of civil wars." *Economics of Governance* 11, no. 2: 123-143.

Moretti, Enrico, and Matthew Neidell. 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources* 46, no. 1: 154-175.

Neidell, Matthew. 2009. "Information, Avoidance Behavior, and Health: the Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources* 44, no. 2: 450-478.

Neidell, Matthew J. 2004. "Air Pollution, Health, and Socio-economic Status: the Effect of Outdoor Air Quality on Childhood Asthma." *Journal of Health Economics* 23, no. 6: 1209-1236.

New York City Independent Budget Office. 2013. "Understanding New York City's Budge: A Guide." Retrieved from http://www.ibo.nyc.ny.us/iboreports/understandingthebudget.pdf.

Payne, A. Abigail, and Justin Smith. 2015. "Does Income Inequality Increase Charitable Giving?" *Canadian Journal of Economics* 48, no. 2: 793-818.

Pickett, Kate E., and Richard G. Wilkinson. 2015. "Income Inequality and Health: A Causal

Review." *Social Science & Medicine* 128: 316-326.

Piketty, T., Saez, E., & Zucman, G. (2016). “Distributional National Accounts: Methods and Estimates for the United States.” National Bureau of Economic Research Working Paper No. w22945.

Ponka, A. 1991. “Asthma and Low Level Air Pollution in Helsinki.” *Archives of Environmental Health: An International Journal* 46(5): 262–270.

Putnam, Robert D. 2007. "Diversity and Community in the Twenty-first Century: the 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30, no. 2: 137-174.

Ransom, Michael R., and C. Arden Pope. 1992. "Elementary school absences and PM 10 pollution in Utah Valley." *Environmental Research* 58, no. 1: 204-219.

Ransom, Michael R. 1995. "External health costs of a steel mill." *Contemporary Economic Policy* 13, no. 2: 86-97.

Reynal-Querol, Marta, and Jose G. Montalvo. 2005. "Ethnic Polarization, Potential Conflict and Civil War." *American Economic Review* 95, no. 3: 796-816.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *The American Economic Review* 94, no. 2: 247-252.

Rohde, Robert A., and Richard A. Muller. 2015. "Air Pollution in China: Mapping of Concentrations and Sources." *PloS one* 10, no. 8.

School Children: A Prospective Cohort Study." *PLoS Med* 12, no. 3.

Roth, Alvin E., and Marilda Sotomayor. 1992. "Two-sided matching." *Handbook of Game Theory with Economic Applications* 1: 485-541., and teacher mobility." *Economics of Education Review* 26, no. 2: 145-159.

Saez, E. 2015. “Striking it Richer: The Evolution of Top Incomes in the United States (Updated with 2012 preliminary estimates).” https://eml.berkeley.edu/~saez/saez-UStopincomes-2014.pdf.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner. 2007. "Race, poverty

Scharber, H., C. Lucier, B. London, A. Rosofsky, and J. Shandra. 2013. “The Consequences of

Exposure to Developmental, Neurological, and Respiratory Toxins for School Performance: a Closer Look at Environmental Ascription in East Baton Rouge, Louisiana.” *Population and Environment* pp. 1–20.

Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83, no. 2: 768-809.

Sunyer, Jordi, Mikel Esnaola, Mar Alvarez-Pedrerol, Joan Forns, Ioar Rivas, Mònica López-Vicente, Elisabet Suades-González et al. 2015. "Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary

Tavoni, Alessandro, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. "Inequality, communication, and the avoidance of disastrous climate change in a public goods game." *Proceedings of the National Academy of Sciences*108, no. 29: 11825-11829.

United States Census Bureau. 2002. Decennial Census 2000: Summary File 3 (SF 3). Retrieved from http://factfinder2.census.gov.

United States Census Bureau. 2010. *2005-2009 American Community Survey 5-year estimates*. Retrieved from http://factfinder2.census.gov.

United States Census Bureau. 2015. *2010-2014 American Community Survey 5-year estimates*. Retrieved from http://factfinder2.census.gov.

Warr, Peter G. 1983. "The Private Provision of a Public Good is Independent of the Distribution of Income." *Economics Letters* 13, no. 2: 207-211.

Wiseman, Jane. 2015. “Can 311 Call Centers Improve Service Delivery? Lessons from New York and Chicago.” *Innovations in Public Service Delivery* 1. Inter-American Development Bank.

Appendices

Table A.1: Policy Changes Affecting I&S from 2009 to 2015

Time	Policy change at all schools
2009-1st round of 2011	1. **External and early-contract candidates**: External candidates only enter after I&S. No early-contracts. 2. **Seniority**: During I&S, schools must interview top 5 most-senior applicants. They then have the option to interview up to 5 other candidates as desired. 3. **Forced excessing**: Any teacher hired into a contract position between February 1 and the start of I&S is "force excessed" so that the vacancy becomes available for I&S. Sites still have the ability to interview the person they had hired and select them for the vacancy but are forced to participate in I&S. 4. **Forced placements**: if internal candidates are excessed due to budget cuts and do not secure a position through I&S, the district places them into whatever open position for which they qualify.
2nd round of 2011-2012	1. **External and early-contract candidates**: Early-contract candidates begin to participate in I&S with greater frequency. They participate in the process with no seniority. 2. **Seniority**: No change. 3. **Forced excessing**: No change. 4. **Forced placement**: No change.
July 2013	1. **External and early-contract candidates:** No change. 2. **Seniority**: During I&S, schools must interview top 4 most senior applicants. Sites then have the option to interview up to 4 other internal and early-contract applicants. 3. **Forced excessing**: Post-Feb 1 hires are not automatically force excessed at priority sites. 4. **Forced placement**: No forced placement of staff at Priority Schools.

Table A.2: Integrity of I&S Rules

Year	Number of applicants within a posting				
	<=5 applicants	**6-10 applicants**		**11+ applicants**	
	Share of postings interviewing all 5 most senior applicants	Share of postings interviewing all 5 most senior applicants	Share of applicants with seniority below cutoff interviewed	Share of postings interviewing all 5 most senior applicants	Share of applicants with seniority below cutoff interviewed
2009	0.99	0.94	0.68	1.00	0.62
2010	0.97	0.90	0.80	0.89	0.49
2011	0.99	0.97	0.84	0.88	0.44
2012	0.97	0.88	0.68	0.98	0.36
2013	0.95	0.91	0.59	0.91	0.35
	<=4 applicants	**5-8 applicants**		**9+ applicants**	
	Share of postings interviewing all 4 most senior applicants	Share of postings interviewing all 4 most senior applicants	Share of applicants with seniority below cutoff interviewed	Share of postings interviewing all 4 most senior applicants	Share of applicants with seniority below cutoff interviewed
2014	0.91	0.85	0.74	0.91	0.31
2015	0.92	0.88	0.69	0.82	0.30

Table A.3: Estimating School Preferences Separately by Stage

	(S3A)	(S3B)	(S4A)	(S4B)
	Resume-screening Stage		Interview Stage	
Age	-0.012678***	-0.012699***	-0.034634***	-0.034611***
	(0.003906)	(0.003807)	(0.002777)	(0.002734)
Female	0.135363	0.139177	0.163556***	0.165877***
	(0.100973)	(0.101046)	(0.056531)	(0.056856)
Teacher of color	-0.255363***	-0.254465***	-0.223474***	-0.219846***
	(0.089979)	(0.090128)	(0.065743)	(0.067663)
Holding an advanced degree	0.258618***	0.260104***	0.115389**	0.119986**
	(0.083229)	(0.083800)	(0.057103)	(0.057173)
Mid-career (4-10 years)	0.181186*	0.197748*	0.087090	0.097859
	(0.104481)	(0.105528)	(0.071488)	(0.071360)
Late-career (over 10 years)	0.675741***	0.667834***	0.032576	0.047909
	(0.131232)	(0.127168)	(0.087667)	(0.087881)
Early offers candidate	0.149979	0.188811	0.248711*	0.267937*
	(0.118485)	(0.118123)	(0.144682)	(0.140355)
Automatic interview	-2.21e+01***	-2.18e+01***	-0.181185***	-0.193052***
	(0.173325)	(0.175698)	(0.065031)	(0.064286)
Average number of years at all previous schools	0.134195***	0.141726***	0.009312	0.015746
	(0.036815)	(0.036330)	(0.016851)	(0.016262)
TE: SOEI (z-score)	0.071434		0.305244***	
	(0.091021)		(0.058867)	
TE: Student survey (z-score)	0.249266***		0.071149	
	(0.096197)		(0.055708)	
TE: Value-added (z-score)	0.000169		-0.122715**	
	(0.106633)		(0.057236)	
TE: Composite (z-score)		0.127125*		0.277073***
		(0.066388)		(0.049044)
Log-likelihood	-2633.8	-2638.7	-5747.6	-5759.9
Chi2	20232.1	18223.9	548.6	399.9
Observations	11,440	11,440	9,899	9,899

Note: In models S3A and S3B, the dependent variable is a binary choice for school *j* whether to send an interview invitation to teacher *i* given that it receives an application from teacher *i*. S3A and S3B are estimated using the conditional mixed logistic regression with position fixed effects. In models S4A and S4B, the dependent variable is an ordered ranking from 1 to 4, where 1 is the highest ranking. S4A and S4B are estimated using the rank-ordered logistic regression. The difference between A and B variants of these models is we break down teacher effectiveness measures into the three sub-components in A and combine them in B. The effectiveness measures are the most recent rolling averages. For more information on the teacher effectiveness measures, see Appendix B. Robust standard errors are clustered at the school level. *p < .10. **p < 0.05. ***p < .01.

Table A.4: Estimating School Preferences using Alternative Ranking

	(S5A)	(S5B)	(S6A)	(S6B)
Age	-0.016293***	-0.016310***	-0.016307***	-0.016323***
	(0.002123)	(0.002117)	(0.002117)	(0.002111)
Female	0.096387**	0.102267**	0.095306**	0.101218**
	(0.046711)	(0.046733)	(0.046628)	(0.046641)
Teacher of color	-0.241498***	-0.239603***	-0.140962	-0.141358
	(0.043327)	(0.043467)	(0.132165)	(0.134589)
Holding an advanced degree	0.167376***	0.163795***	0.167278***	0.163676***
	(0.052908)	(0.053401)	(0.052937)	(0.053434)
Mid-career (4-10 years)	0.114684*	0.122872*	0.115087*	0.123314*
	(0.065334)	(0.066067)	(0.065073)	(0.065779)
Late-career (over 10 years)	0.208567***	0.210248***	0.209802***	0.211448***
	(0.066856)	(0.065951)	(0.066689)	(0.065755)
Early offers candidate	0.125611	0.139175*	0.126000	0.139682*
	(0.080771)	(0.083128)	(0.080750)	(0.083094)
Automatically interview	0.822628***	0.823616***	0.822795***	0.823778***
	(0.053634)	(0.053595)	(0.053595)	(0.053560)
Average number of years at all previous schools	0.092872***	0.098137***	0.092661***	0.097929***
	(0.018314)	(0.017941)	(0.018341)	(0.017965)
TE: SOEI (z-score)	0.122951***		0.123007***	
	(0.045759)		(0.045719)	
TE: Student survey (z-score)	0.119142**		0.119790**	
	(0.049362)		(0.049294)	
TE: Value-added (z-score)	0.049634		0.048744	
	(0.063436)		(0.063172)	
TE: Composite (z-score)		0.155379***		0.155574***
		(0.042798)		(0.042817)
Teacher of color X % student of color at receiving school			-0.001386	-0.001355
			(0.001617)	(0.001660)
Log-likelihood	-8044.5	-8051.3	-8044.2	-8051.0
Chi2	1007.1	896.5	1078.3	942.9
Observations	12,427	12,427	12,427	12,427

Note: The above results are obtained from running rank-ordered logistic regressions where the receiving-schools are the decision makers. The choice sets are the pools of applications. For all models, the dependent variable is an ordered ranking from 1 to 4, where 1 is the highest ranking. Applicants who were screened out before the interview stage receive a censored ranking of zero. Automatic interview applicants who were not ranked in the top are assigned the maximum ranking within the pool plus one. If there is more than one such applicant, they all will share this rank. The difference between A and B variants of these models is we break down teacher effectiveness measures into the three sub-components in A and combine them in B. The effectiveness measures are the most recent rolling averages from the first year of work up to the year of application. For more information on the teacher effectiveness measures, see Appendix B. Robust standard errors are clustered at the school level. *p < .10. **p < 0.05. ***p < .01.

Figure A.1: Total Number of Applications by Year

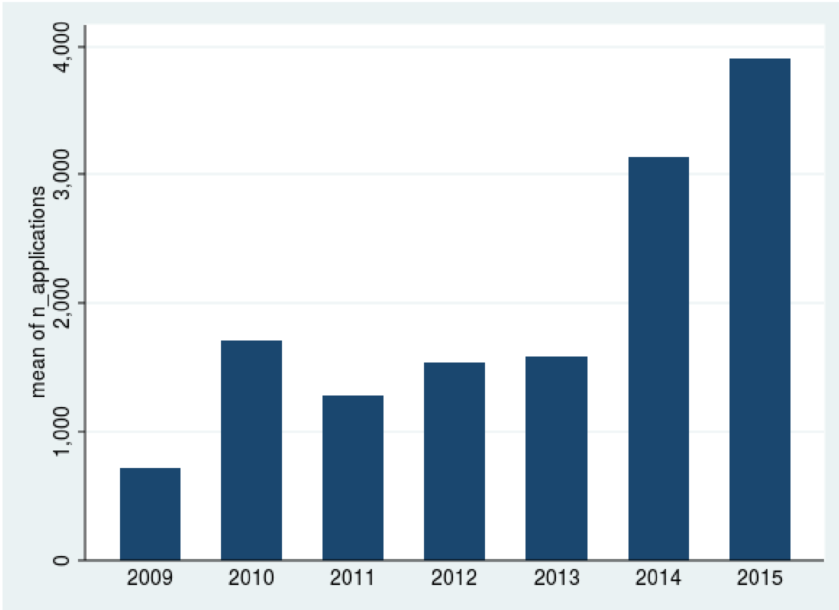


Figure A.2: Total Number of Unique Applicants by Year

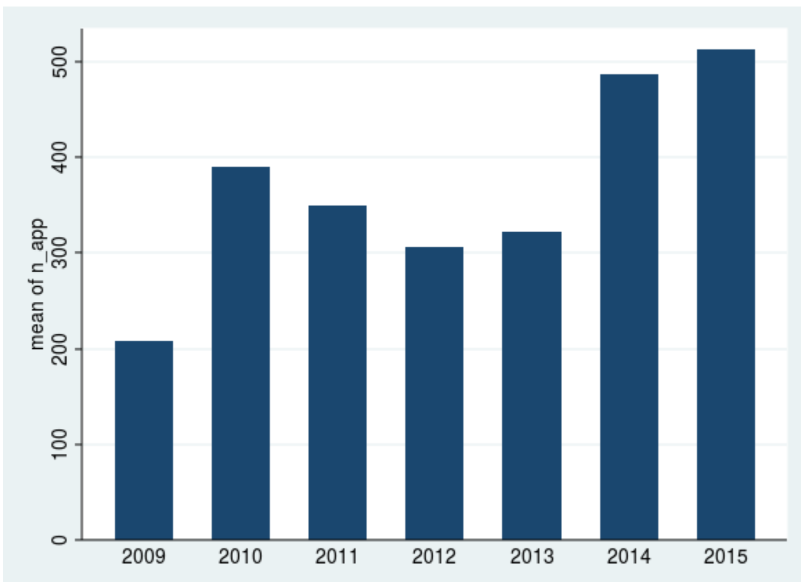


Figure A.3: Number of Postings by Year

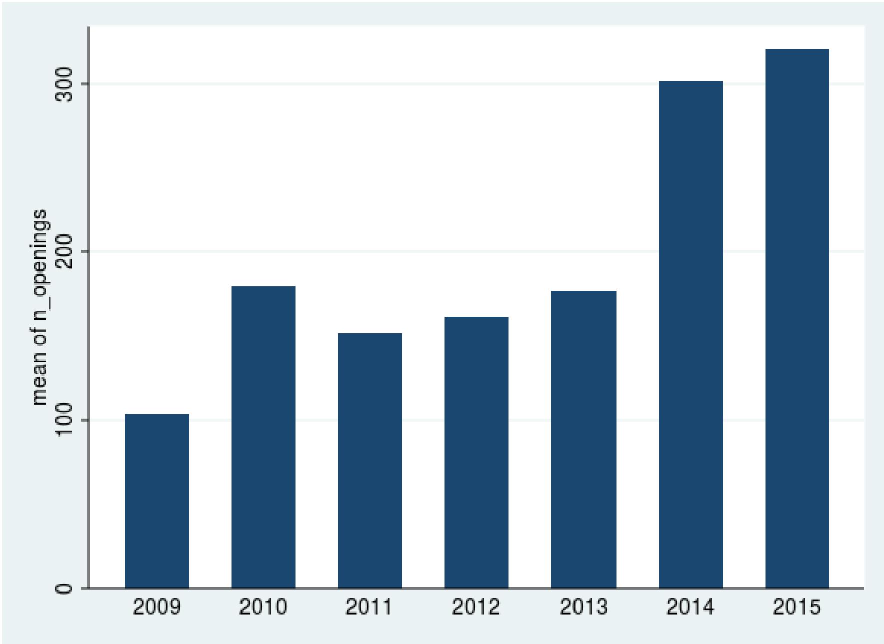
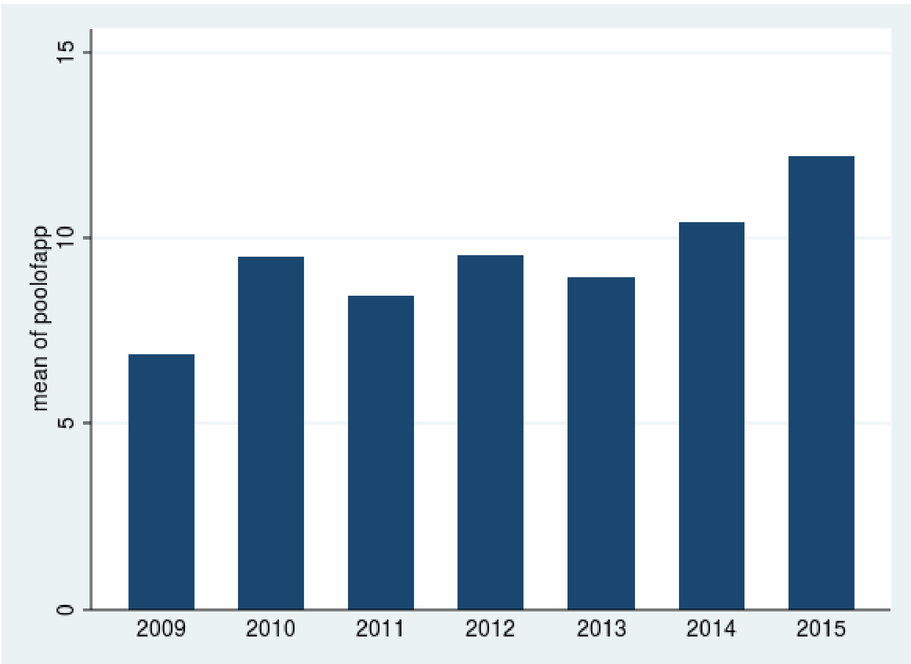


Figure A.4: Average Size of Application Pool by Year



Appendix B:

The district uses these measures to operationalize teacher effectiveness as three first-order factors measured at different times: 1) sound instructional practices (SOEI), 2) student engagement and perspectives of instruction, and 3) value added by a teacher to student academic growth. The value-added scores use teacher-verified rosters and measure the average contribution to student test score growth. This calculation uses modeling that accounts for pre-test achievement and other student characteristics are exogenous sources of construct-irrelevant variance in post-test scores.[29] In short, the model isolates the progress each teacher's students make on state mandated standardized tests relative to other students who had similar pre-test scores. The student surveys are a version of the surveys designed by Ronald Ferguson (cite).[30] The survey is administered twice each year and ask K-12 students about the degree to which their teachers academically "engage" (3-8 items depending on grade level), "illuminate" (3-7 items), "manage" (3-6 items), "relate" (3-7 items), and "stretch" (3-7 items) them and their peers. The formal observations are conducted by a supervisor or other trained raters and use version of the rubric developed by Charlotte Danielson to measure effective teaching (cite).[31] The rubric specifies sound instructional practices in terms of three domains: planning and preparation (seven items), classroom environment (five items), and (nine items).

---

[29] The overall reliability of value-added scores varies considerably by grade from about 0.75-0.90 for math and 0.44- 0.86 for reading. The reliability/standard error of measurement of an individual teacher's score depends on the number of students taught and the proportion of variation in post-test scores explained by the control variables.

[30] 2 Student surveys are scored using generalizability theory. Generalizability theory is appropriate because the design is balanced and students qualify as rater facets (but for whom reliable estimates of rater effects are not possible, in contrast to observer effects). The overall reliability of student survey scores is about 0.81 in grades K-2, 0.82 in 3-5, and 0.90 in 6-12. The reliability/standard error of measurement of an individual teacher's score depends on the number of students and items.

[31] Classroom observations are scored using the many-facets Rasch rating scale model. The many-facets Rasch model is used because it accommodates the highly unbalanced design (i.e., missingness due to short observations) and adjusts for variation in item difficulty and rater severity. Additionally, estimates of rater effects enable us to identify extreme and/or misfitting observers who need additional training. The overall reliability of observation scores is about 0.87. The reliability/standard error of measurement of an individual teacher's score depends on the number of observations, raters and items.

Appendix C:

Derivation of Marginal Rates of Substitution

$$U_i = \boldsymbol{\beta}(\boldsymbol{X}_{ky} - \boldsymbol{X}_{jy}) + \epsilon_{ijky}$$

$$U_i = \boldsymbol{\beta}\boldsymbol{D}_{jky} + \epsilon_{ijky}$$

$$dU_i = (\partial U/\partial D_{1,jky})dD_{1,jky} + \cdots + (\partial U/\partial D_{m,jky})dD_{m,jky}$$

$$dU_i/dD_{1,jky} = (\partial U/\partial D_{1,jky}) + (\partial U/\partial D_{2,jky})\frac{dD_{2,jky}}{dD_{1,jky}} + \cdots + (\partial U/\partial D_{m,jky})\frac{dD_{m,jky}}{dD_{1,jky}}$$

$$0 = MU_{D1} + MU_{D2}\frac{dD_{2,jky}}{dD_{1,jky}} + \cdots + MU_{Dm}\frac{dD_{m,jky}}{dD_{1,jky}}$$

If we allow only attributes 1 and 2 to change ($dD_{n,jky}$ = 0 for all $n > 2$)

$$-\frac{MU_{D1}}{MU_{D2}} = \frac{dD_{2,jky}}{dD_{1,jky}}$$

$$MRS_{D1,D2} = \frac{MU_{D1}}{MU_{D2}}$$

$$MRS_{D1,D2} = \frac{\beta_1}{\beta_2}$$

$MRS_{D1,D2}$ is the rate at which a teacher is ready to give up some amount of school attribute 1 in exchange for some amount of school attribute 2 while maintaining the same level of utility.